# 目　　录

## 第三篇 比较静态分析







## 第四篇　最优化问题

## 第五篇　动态分析

## 第六篇 数学规划

# 序　言

本书是为那些致力于学习基本数学方法的经济学专业的学生而写的。这些数学方法已成为正确理解当前经济文献的必不可少的工具。不幸的是，对于多数学生而言，学习数学恰如服食苦口良药：必要而且难以回避，却又非常艰涩。这种被称之为“数学焦虑症”的态度的根源，我相信很大程度上在于数学不当的陈述方式。在“简洁即美”的信念的驱使下，数学解释有时只是为了清晰而过于简洁，使学生感到困惑，产生一种认为自身知识不足的错误意识。过分正式的表述，如果不伴以直观的解释，或者“贴切”的证明，就会挫伤学生学习的动力；内容水平的参差不齐也会使某些数学问题表现得比它们实际的难度要大。最后，练习题的过分复杂，往往会击垮学生的信心，而不是激发他们思考。

基于以上认识，我作了极大努力以缩小导致数学焦虑症的问题。在可能的范围内，我对有关内容进行了耐心而又明晰的解释，并审慎地采用了一种非正式的“用户至上”(这是从计算机行业借来的术语)的方式。在顺序的安排上，我力图预测并且回答那些学生在阅读时，心中有可能提出的问题。为强调数学对经济学的价值，我让经济学家的分析需要驱动对相关数学方法的学习，然后立即以适当的经济模型图示说明这种数学方法。而且，一系列数学工具的介绍也是认真地建立在循序渐进的基础之上的。我首先介绍基本的数学工具以作为后面要讨论的更复杂的工具的阶梯。在

合适的场合，还采用图形解释以对代数的结果予以直观的支持。此外，我们设计了一些练习题(更多的是作为小练习)。这些练习题更能帮助巩固掌握的知识，并激发自信心，而不是施加那种有可能不小心地挫伤并吓跑初学者的挑战。

本书涵盖如下主要经济分析的内容：静态学(均衡分析)、比较静态学、最优化问题(静态学的一种特例)、动态学和数学规划(最优化的现代发展)。为掌握上述内容，我介绍了如下数学方法：矩阵代数、微积分、微分方程、差分方程和凸集。由于书中介绍了大量宏观、微观经济模型，所以，本书对那些已受过数学训练，但需要一个向导，引导其由数学王国步入经济学殿堂的人来说，也是极有裨益的。基于同样的原因，本书不仅可以作为数学方法的教科书，而且也可以作为学习宏观经济理论、微观经济理论、经济增长与经济发展理论等课程的补充读物。

尽管依然保持了前两版的主要目的、风格、结构，本版还是作了如下几个重要修改：介绍最优化的古典方法的第 11 章和第 12 章已被大量重写。我采用一阶、二阶条件的微分(与导数相对照)形式作为贯串这两章的统一主题，尽管一阶、二阶条件的导数形式仍作为解题的运算准则。更重要的是，我增加了对与自由极值问题的二阶条件相关的凹凸性问题(11.5 节)、与等式约束的最优化问题(12.4 节)相关的拟凹、拟凸性问题的深入讨论。这种安排不仅便于理解二阶条件，而且也使得更早地介绍绝对极值的内容成为可能，通常，绝对极值的内容是不在古典方法中强调的。这两方面的讨论概括在两个新图——图 11.5 和图 12.6 中。

另一个主要变化体现在对联立微分方程组和联立差分方程组的处理方面。在本版增加了新的一章(第 18 章)中，进一步讨论了

上述课题,这包括增加了对双变量相位图方法(18.5 节)、非线性微分方程组的线性化与局部稳定性分析(18.6 节)等内容的介绍。

另外,本版还有其它改进:增加了对齐次函数的简短讨论(12.7 节);简化了对连续性与可微性的解释(6.7 节);对二阶必要条件予以更多的强调(9.4 节、11.1 节和 12.3 节);澄清了对将“泰勒展开”(Taylor expansion)作为任意函数的近似的解释(9.5 节);至于经济解释方面,我增加了通货膨胀与失业相互作用模型作为微分方程与差分方程应用的例子,这个模型还涉及诸如预期增加的菲利蒲斯关系和适应性预期等概念的运用(15.5 节、17.3 节、18.4 节)。应读者要求,本书后面给出了部分习题的答案;为便于参考,我在书后增加了一个数学符号附录;为尽可能容纳更多的新内容,我极不情愿地删去了博弈论一章。

在写作本书的过程中,我应向许多人表示感谢。特别是应当向那些数学家和经济学家表示感谢。本书借鉴了他们的许多思想,却又未曾提及其名字。对于本书的前两版,我得益于以下人士的评论和建议(以英文字母顺序排列):南希·S. 巴雷特、托马斯·伯恩伯格、E. J. R. 布思、罗伯塔·格罗尔·凯里、埃米莉·蒋、劳埃得·R. 科恩、哈拉尔德·迪克森、约翰·C. H. 费、罗杰·N. 福尔瑟姆、杰克·伊尔茨雷费尔、詹姆斯·肖、荣凯军、马克·纳洛夫、J·弗兰克·夏普、丹尼斯·斯塔利夫以及叶周南等。对于本版,下列人士也提出了许多极有价值的建议,他们是:E. J. R. 布思、查尔斯·E. 巴特勒、加里·康纳尔、沃伦·L. 费希尔、丹尼斯·R. 赫夫利、乔治·孔德尔、威廉·F. 洛特、保罗·B. 曼彻斯特、彼特·摩根、阿伦·G. 斯利曼以及最后但不是不重要的亨利·Y. 万等。我谨向他们表示诚挚的谢意,但因为我并未接受他们的全部建议,所以,本书的责任依然

由我独立承担。特别是我再一次决定,不把动态最优化的若干章节包括在本书中。这部分内容,我将在另一本书中详尽探讨。最后,我还要向麦克劳出版公司的盖尔·戈韦特表示衷心的感谢,感谢她耐心的合作以及对复杂的原稿的熟练的处理技巧。

**对如何使用本书的建议**

由于本书的内容是按数学工具循序渐进的演化顺序来组织的,所以学习本书的理想方式当然是按本书的具体展开顺序来进行。但有一点重要的例外:先学习第五篇(动态分析)还是第六篇(数学规划)是无关紧要的,不把比较静态学作为主要研究领域的读者,可以跳过一般函数模型的比较静态分析(第 8 章)这部分内容,从第 7 章直接过度到第 9 章。但在这种情况下,这些读者也要省略 11.7 节和 12.5 节中的比较静态分析部分。

蒋 中 一

# 第一篇

# 导　　论

# 第1章　数理经济学的实质

公共财政和国际贸易是经济学的分支学科。从这个意义上看，数理经济学显然不是经济学的一个分支学科。确切地说，它是一种经济分析方法，是经济学家利用数学符号描述经济问题，运用已知的数学定理进行推理的一种方法。就分析的具体对象而言，它可以是微观或宏观经济理论，也可以是公共财政、城市经济学，或者其它经济学科。

从更广泛的意义上使用数理经济学这一概念，我们甚至可以将现在的任何一本初级经济学教程均称之为数理经济学，因为它们经常运用几何学方法推导理论结果，然而，习惯上，数理经济学是指不仅使用简单的几何学方法，而且还运用像矩阵代数、微积分、微分方程、差分方程等数学工具来描述经济问题的一种方法。本书的目的就是向读者介绍这些数学方法的基本内容。这些内容是在现代经济学文献中经常遇到的。

## 1.1　数理经济学与非数理经济学

因为数理经济学仅仅是一种经济分析的方法，所以，它与非数理的经济分析方法不应当，事实上也不存在任何根本的不同。任何一种理论分析，不管运用何种方法，其目的总是从一些给定的假设或公理出发，通过推理过程得出一组结论或定理。数理经济学

与所谓“文字经济学”(Literary economics)的主要区别基本源于以下事实:前者使用数学符号而非文字、使用方程而非语句来描述假设和结论;其次,前者运用大量的可供引用的数学定理而非文字逻辑进行推理。因为符号和文字表述实际上是相同的(符号通常用文字加以定义便可证明这一点),所以,选择哪一种表述方式并无实质差别。不过,人们公认的是,数学符号更便于演绎推理,且能使表述更为言简意赅。

另外,选择文字逻辑和数学逻辑虽然并无实质差别,但运用数学推理有这样一个优势:它可以促使分析者在推理的每一阶段都作出明确的假设,这是因为数学定理通常是按“如果—那么”形式加以陈述的,所以,为了导出所运用定理的“那么”(结论)部分,分析者必须确保“如果”(条件)部分与其所采纳的明晰假设相一致。

纵使假设上述观点都正确,可能有人仍会问:有何必要超越几何学方法呢? 答案是,尽管几何分析具有直观性这一重要优点,但它也苦于严格的维数限制。例如,在无差异曲线的一般图形讨论时,标准的假设是消费者仅能得到两种商品,采用这种简化的假设并非出于自愿,而是不得已而为之,因为绘出三维空间的几何图形极为困难,要绘出四维(或更多维)空间的几何图形是根本不可能的。要研究三种、四种或几种商品等更为一般的情况,我们必须寻求更为灵活的工具——方程。仅此一个理由已足以使我们研究几何学以外的数学方法。

简而言之,我们看到数学方法具有如下优点:(1)所运用的“语言”更为简练精确;(2)有大量的数学定理可为我们所用;(3)它迫使我们明确陈述所有假设,作为运用数学定理的先决条件;这能使我们戒除不自觉地采用不明确的假设的缺点;(4)使我们能够处理

$n$ 个变量的一般情况。

与这些优点相对照，人们也能听到这种批评：以数学推导的理论必然是不现实的。然而这种批评并不切中要害。事实上，“不现实的”这种说法甚至也不能用于批评一般的经济理论，无论其所运用的方法是数理性的还是非数理性的。就其本质而言，理论是对现实世界的抽象，是一种找出最重要的因素和联系，以使我们对所关注问题的核心进行研究的一种手段。它能使我们免于陷入现实世界中确实存在的种种复杂性而难以自拔。因此，“理论缺乏现实性”这种批评只不过是一种陈词滥调，不能视为对理论的一种有效的批评。从逻辑上看，把任何一种探讨理论的方法都视为“不现实的”方法是极其无意义的。例如，完全竞争的厂商理论与不完全竞争的厂商理论一样，都是不现实的，但这些理论是否用数学推导，与此无关，也并不重要。

总之，我们可以把数学方法视为一种“运输方式”，它能使我们以最快的速度从一组假设（出发点）出发，得到一组结论（目的地）。常识告诉我们，如果你想去两英里远的地方，你会宁愿开车而不愿步行，除非你想消磨时间或想锻炼身体。类似地，作为一个希望尽快得出结论的理论家，驾驶数学工具这一“机动车”，更便于实现其特殊目的。当然，你必须首先学习“驾驶课程”；但由于掌握这些技能具有长期的价值，适当投入一些时间和精力是值得的。

对于一个严肃的“司机”来说——我们继续沿用上述比喻——具有牢固的数学基础是绝对必要的。很显然，在这样一部著作中介绍清楚经济学家所运用的全部数学工具是不可能的。因此，我将集中介绍那些数学上最为基础、与经济学最为相关的内容。尽管如此，如果你认真学完本书，你至少会非常熟练地理解

偶然碰到的下列经济学期刊中的大部分专业论文：《美国经济评论》、《经济学季刊》、《政治经济学杂志》、《经济学与统计学评论》以及《经济学杂志》等。学习了上述内容之后那些对数理经济学产生了极大兴趣的人，则可以进行更严格的数学训练，学习更高深的数学内容。

## 1.2 数理经济学与经济计量学

“数理经济学”这一术语有时常与另一个与其相关的术语“经济计量学”相混淆。正如这术语的后一部分“计量学”所表明的那样，经济计量学主要与经济数据的度量有关，它运用估计和假设检验的统计学方法进行经验观测的研究。而数理经济学则是把数学应用于经济分析的纯理论方面，基本不涉及或不关心诸如所研究的变量的度量误差这类统计问题。

本书仅限于研究数理经济学，即主要集中于将数学应用于演绎推理而非归纳研究，因而，我们将主要进行理论研究而非经济分析。当然，这仅仅是一个研究范围的选择问题，决不意味着经济计量学不如数理经济学重要。

实际上，经验研究和理论分析是相辅相成、相互促进的。一方面，理论在有把握地应用之前，必须运用经验数据对其有效性进行检验。另一方面，要确定关系最为密切和最富有成效的研究方向，
6
统计工作必须有理论作为指南，理论研究和经验研究互补关系的经典描述在总消费函数的研究中可以找到。凯恩斯关于消费函数的理论研究导致对消费倾向的统计估算，而库兹涅茨和戈德史密斯关于消费倾向相对长期的稳定性(与凯恩斯理论所预期的结果

相矛盾)，又促进了杜森贝里、弗里德曼等人对总消费理论的完善。[1]

然而，从某种意义上说，数理经济学在这两者之中更具基础性：因为要进行有价值的统计和经济计量研究，有一个好的理论框架——最好是数学公式——是必不可少的。所以，本书的内容不仅对热衷于理论经济学的人有所帮助，而且对那些研究经济计量学而又缺乏数学基础的人也极有价值。

---

① 约翰·M.凯恩斯：《就业利息和货币通论》，哈尔科特、布赖斯出版公司，纽约，1936年版，第三卷。西蒙·库兹涅茨：《国民收入：发现的总结》，国家经济研究总署，1946，p 53。雷蒙·戈德史密斯：《美国储蓄研究》，第一卷；普林斯顿大学出版社，普林斯顿，N.J.1955，第三章。詹姆斯·S.杜森贝里：《收入、储蓄和消费者行为理论》，哈佛大学出版社，剑桥，马萨诸塞州。米尔顿·弗里德曼：《消费函数理论》，国家经济研究总署，普林斯顿大学出版社，N.J.，1957。

# 第2章　经济模型

7

如前所述，任何经济理论都是对现实世界的必要抽象。首先，由于现实经济的极端复杂性，我们不可能一下子理解其全部内在联系；其次，这些内部关系对于理解我们所研究的特殊经济现象，也并不具有同样的重要性。因此，合理的研究程序应当是：根据我们的目的选择与我们研究问题相关的基本因素和基本关系，然后把我们的研究集中于这些因素和关系上。这种精心简化的分析结构被称作经济模型，因为它只是现实经济的结构性的粗略性的表示。

## 2.1　数学模型的构成

经济模型仅仅是一种理论框架，并没有必须采用数学形式的内在理由。然而，如果模型是数学模型，那么，它通常包括一组用以描述模型结构的方程。这些方程以某种方式把一定数量的变量联系在一起，并给出所采用的一组分析假设的数学形式。然后通过对这些方程进行相应的数学计算，我们便可以推导出一系列在逻辑上服从这些假设的结论。

### 变量、常数和参数

8

变量是大小可以变化的量，即可以取不同值的量。经济学中

经常使用的变量包括价格、利润、收益、成本、国民收入、消费、投资、进口、出口等等。因为这些变量可以取不同的值，所以必须用一个符号而不是特定的数字表示它。例如，我们可以用 $P$ 表示价格，以 $\pi$ 表示利润，以 $R$ 表示收益，以 $C$ 表示成本，以 $Y$ 表示国民收入等等。但是当我们写出 $P=3$ 或 $C=18$ 时(选择适当单位)，我们就把这些变量"固定"在这些具体的数值上。

通过解一个适当构建的经济模型，我们可以得到一组变量的**解值**，如市场出清时的价格水平，利润最大化时的产出水平等。其解值可以通过模型求出的变量，称作**内生变量**(源于模型内部)，但模型中也包含一些由模型外部因素所决定的变量，其大小仅被视为给定的数据，这样的变量被称为外生变量(源于模型之外)。需要指出的是，一个模型中的内生变量可能是另一个模型的外生变量。例如，在分析小麦市场价格的决定时，变量 $P$ 无疑是内生的；但在消费者支出理论的分析架构内，$P$ 对于个别消费者而言是一个常数，因此必然被视为外生的。

变量常常与固定的值或常数结合出现，像表达式 $7P$ 或 $0.5R$。常数是一个大小不变的值，所以它恰好与变量相反。当一个常数与一个变量结合在一起时，我们将其称作该变量的**系数**。但系数也可以是符号而非数字。例如，为了获得更广泛的一般性(参见 2.7 节)，在模型中我们可以用 $a$ 代表给定常数，以表达式 $aP$ 代表 $7P$。这个符号 $a$ 极为特殊，——我们假定它表示给定常数，但因我们未赋予其具体的数值，它实际上可以取任意值。简而言之，它是一个**可变的常数**！为便于识别其特殊身份，我们将其明确命名为**参常数**(或简称为**参数**)。

必须强调指出，尽管参数可取不同的值，但在模型中仍要将其

视为已知数。正是由于这个原因，即使常数是一个参数，人们还是将其简称为“常数”。在这方面，参数与外生变量极为相似，因为二者在模型中均被视为“给定值”。这可以解释为什么许多作者为了简便，将二者用一个术语“参数”来表示。

按照习惯，参常数通常以字母 $a$、$b$、$c$，或者相应的希腊字母 $\alpha$、$\beta$、$\gamma$ 表示。但也可以使用其它符号表示，至于外生变量，为使其与内生变量明显区别开来，我们用对选定的符号加下标 0 的方式来表示。如 $P$ 表示价格，则 $P_0$ 表示外生的价格。

## 方程和恒等式

9

变量固然可以独立存在，但只有通过方程或不等式将其联系起来才具有实际价值。在这里，我们只讨论方程。

在经济学的应用范围内，我们需要区别三种类型的方程：即定义方程、行为方程和均衡条件。

**定义方程**是在两个具有完全相同含义的不同表达式之间建立恒等式。对于这种方程，我们通常用恒等符号“≡”(读作“恒等于”)代替通常的等号“=”，尽管后者也是可以接受的。比如，总利润被定义为总收益与总成本之差，因此我们可以写成：

$$\pi \equiv R - C$$

而**行为方程**则规定了当其它变量变化时，某一变量相应的变化方式。这可能包括人类行为，比如当国民收入变化时总消费模式的变化；也可能不包括人类行为，比如厂商的总成本如何随产出的变化而变化。从广义上讲，行为方程可以用于描述一般的制度性模型，包括技术方面(如生产函数)的模型，也包括法律方面的模型(如税收结构)。但在写出行为方程之前，必须对所研究的变量

的行为模式作出明确的假设。考察下面两个成本函数

$$C = 75 + 10Q \tag{2.1}$$

$$C = 110 + Q^2 \tag{2.2}$$

其中 $Q$ 表示产量。由于两个方程具有不同的形式，所以每一方程所假设的生产条件必然与另一个方程有明显不同。在(2.1)式中，固定成本为 75(当 $Q=0$ 时的 $C$ 值)，而(2.2)式则为 110。成本的变化也不同。在(2.1)式中，产量 $Q$ 每增加一个单位，$C$ 固定增加 10 单位。而在(2.2)式中，当 $Q$ 一个单位一个单位增加时，$C$ 的增加越来越大。很明显，正是通过对行为方程形式的设定，我们才给出模型所采纳的假设的数学表达式。

第三类方程，**均衡条件**，只有当我们的模型包含均衡这一概念时才会涉及。如果模型中包含这一概念，则均衡条件就是描述实现均衡前提条件的方程。经济学中人们最为熟悉的两个前提条件是：

$Q_d = Q_s$ ［需求量 = 供给量］

和 $S = I$ ［合意储蓄 = 合意投资］

它们分别与市场均衡模型和最简单的国民收入均衡模型有关。因 10 为这类方程既非定义方程，也非行为方程，所以它们单独构成一类方程。

## 2.2 实数系

方程和变量是数学模型的基本组成部分，但因经济变量的取值一般为数字，因此我们需要简单介绍一下数系的概念。这里我们只讨论所谓的“实数”。

像1、2、3……这类整数被称作正整数。正整数最常用于计数。其对应的负值，如 -1，-2，-3，……称作负整数，例如，可以用其表示零度以下的温度(单位为度)。而数字0，既非正数，也非负数，在此意义上看，它是唯一的。我们把所有的正整数、负整数以及0归为一类，称其为整数集合。

当然，整数并未穷尽所有可能的数，因为我们还有诸如2/3，5/4，7/3之类的分数，如果将其标在直尺上，它们位于两个整数之间。同样，我们还有负分数，如 -1/2，-2/5 等。正负分数合在一起，构成了分数集合。

所有分数的一个共性是可以将其表示为两个整数之比。这样，可把分数称之为有理数(在这里"有理"意指比值)。但任何一个整数都是有理数，因为任意整数 $n$，均可视为比率 $n/1$。整数集合与分数集合合在一起，构成了有理数集合。

当然，我们一旦运用了有理数这个概念，自然会引出无理数这个概念。无理数是不能表示为两个整数之比的数。如 $\sqrt{2} = 1.4142\cdots\cdots$，它是一个无限不循环小数；再如 $\pi = 3.1415\cdots\cdots$(表示圆的周长与直径之比)，也是一个无限不循环小数，其小数部分为无限不循环的，这是所有无理数的特征。

如果将一个无理数标在尺上，则它将居于两个有理数之间。正如分数可以填满两个整数之间的空隙一样，一些无理数也可以填满两个有理数之间的空隙。这种填充的过程会产生一个数的连续统，它被称作"实数"。连续统构成全部实数集合，常以符号 $R$ 表示。当我们将集合 $R$ 标示在直线上时(一个加长的尺子)，我们将其称作实线。

在图2.1中，我们根据讨论顺序及彼此之间的相互关系，列出

了所有的数集。但如果我们从下向上看这个图,实际上我们得到一个分类图解:即实数集可以分解成它的子集以及子集的子集。所以这个图是实数系结构的一个概览。

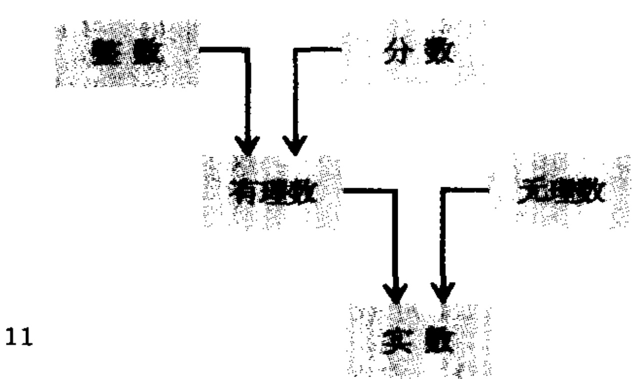



图 2.1

11

实数已能满足本书前 14 章的需要,但它并非数学中使用的全部的数。事实上,使用“实数”这个概念就意味着有“虚数”的存在,虚数与负数的平方根有关,我们将在第 15 章中讨论。

## 2.3 集合的概念

我们已多次使用“集合”一词。由于集合这一概念已成为现代数学各分支的基础,我们至少要掌握其基本内容,才能满足我们的需要

### 集合的符号

简言之,一个**集合**就是不同对象的集成。这些对象可以是不同的数,也可是别的什么东西。因此,如同三个整数 2,3,4 可以构成一个集合一样,所有选修一门特定经济学课程的学生也可以视为一个集合。集合中的对象称作集合的**元素**。

有两种书写集合的方法:**列举法**和**描述法**。如果令 $S$ 表示由 2,3,4 三个数组成的集合,用元素列举法可以写成:

$S=\{2,3,4\}$

但若令 $I$ 代表所有正整数的集合，列举便很困难，则我们可以简单地描述其元素，并写成：

$I=\{x \mid x$ 为一任意正整数$\}$

此式可读作："$I$ 为满足 $x$ 为一个正整数的所有 $x$ 的集合。"注意，集合两边均使用大括号将其括住。使用描述法，一定要在代表元素的一般符号和所描述的元素之间加一竖杠或冒号，以便将二者区分开来。再举一例，大于 2 小于 5 的实数集合 $J$ 可以用符号表示为

12

$$J=\{x \mid 2<x<5\}$$

这里，描述性陈述也是以符号表示的。

一个具有有限元素的集合，如上例中的集合 $S$，称为**有限集合**。而集合 $I$ 和 $J$ 均有无限个元素，称为**无限集合**。有限集合总是可数（或可计算）的，即其元素数量可以以基数 1，2，3，……表示。但无限集合既可能是可数的，如集合 $I$，也可能是不可数的，如集合 $J$。我们无法将后者中的元素与基数 1，2，3，……等联系起来，所以是不可数的。

集合中的成员以符号 $\in$（希腊字母 $\epsilon$ 即 *epsilon* 的变形，表示"元素"）表示，读作"属于…的元素"。因此，对于我们前面定义的集合 $I$ 和 $J$，我们可以写出

$2\in S, 3\in S, 8\in I, 9\in I$ 等

但显然 $8\notin S$（读作：8 不是集合 $S$ 的元素）。如果我们使用符号 $R$ 表示所有实数集合，则"$x$ 是某些实数"可简单表示成

$$x \in R$$

## 集合间的关系

将两个集合加以比较时，可能观测到几种可能的关系。若两

个集合 $S_1$ 和 $S_2$ 恰好有相同的元素：

$$S_1 = \{2,7,a,f\} \text{ 和 } S_2 = \{2,a,7,f\}$$

则称 $S_1$ 和 $S_2$ 相等($S_1 = S_2$)。注意集合中元素出现的次序是无关紧要的。但只要其中有一个元素不同，两个元素便不相等。

另一种关系是，一个集合可能是另一集合的子集。若我们有两个集合

$$S = \{1,3,5,7,9\} \text{ 和 } T = \{3,7\}$$

则 $T$ 是 $S$ 的子集，因为集合 $T$ 的每一个元素都同时是集合 $S$ 的元素。对这种情况更为正式的表述为："当且仅当'$x \in T$'意味着'$x \in S$'时，$T$ 为 $S$ 的子集"。使用集合符号 $\subset$(包含于)和 $\supset$(包含)，则我们可以写成

$$T \subset S \text{ 或 } S \supset T$$

两个集合恰好互为子集也是可能的。但这必然意味着两个集合相等。我们可以将此正式表述为：当且仅当 $S_1 = S_2$ 时，我们可以有 $S_1 \subset S_2$ 且 $S_2 \subset S_1$。

13 需要注意的是，符号 $\in$ 表示个别元素与集合之间的关系，而符号 $\subset$ 则表示子集与集合的关系。我们可以运用这一思想对图 2.1 中的某些关系进行描述：整数集合是有理数集合的子集；同样，无理数集合是实数集合的子集。

具有 5 个元素的集合 $S = \{1,3,5,7,9\}$ 可以形成多少子集呢？首先，$S$ 中的每一个元素，可单独构成一个 $S$ 的子集，如 $\{1\}$，$\{3\}$ 等。同样，$S$ 中的任意两个，三个、四个元素，均可构成 $S$ 的子集，如 $\{1,3\}$，$\{1,5\}$，…，$\{3,7,9\}$，等等。而且，集合 $S$ 本身(包括全部 5 个元素)也可视为其自身的子集——$S$ 的每一个元素均是 $S$ 的元素，因此它符合子集的定义。当然，这是一种极端的情况，它使

我们得到集合 $S$“最大的”子集，即 $S$ 自身。

另一种极端的情况是 $S$ 的“最小的”可能的子集，即它本身不包含任何元素。这种集合被称作零集，或者空集，以符号 $\varnothing$ 或 $\{\}$ 表示。将零集称作 $S$ 的子集的理由非常有趣：如果零集并非 $S$ 的子集（$\varnothing \not\subset S$），则 $\varnothing$ 必定至少包含一个元素使得 $x \notin S$，但根据定义，零集不含有任何元素，所以我们不能说 $\varnothing \not\subset S$；因此零集为 $S$ 的子集。 14

计算 $S$ 的所有子集，包括两种极端情况 $S$ 和 $\varnothing$，我们知道共有 $2^5 = 32$ 个子集。一般而言，若一个集合含有 $n$ 个元素，则它共可形成 $2^n$ 个子集。[①]

将符号 $\varnothing$ 或 $\{\}$ 明确地与 $\{0\}$ 区别开来是至关重要的。前者不含有任何元素，而后者则包含一个元素 0。空集是惟一的，全世界只此一个集合，它可以是我们可以构想出来的任何一个集合的子集。

第三种可能的关系类型是：两个集合不含有任何共同的元素。在这种情况下，我们称两个集合不相交。如所有正整数的集合与所有负整数的集合是不相交的集合。第四种关系类型是：两个集合具有某些共同的元素，但另一些元素分属不同的集合。在此情况下，两个集合既不相等，也并非不相交；且每一集合也不是另一个集合的子集。

---

① 给定一个具有 $n$ 个元素的集合 $\{a, b, c, \cdots, n\}$，我们首先可将其子集分为两类：一类包括元素 $a$，另一类不包括 $a$。而上述每一类还可以进一步分成两个子类：一类含有元素 $b$ 而另一类则没有；注意，我们在考察第二个元素 $b$ 时，我们已把类别的数量增加了一倍，即从 2 增加到 $4(=2^2)$。同样，考察元素 $c$，将使类别总数增至 $8(=2^3)$。当考察到 $n$ 个元素时，类别的总数将等于子集的总数，其数量为 $2^n$。

## 集合的运算

当我们对某些数进行加、减、乘、除或开方时，我们便是在对其进行数学运算。集合不同于数，但我们仍可对其进行类似的数学运算。这里讨论三种基本的运算，即并集、交集和补集。

取两个集合的并集就是构成一个新的包含且仅包含属于集合 $A$、$B$ 或者同时属于 $A$ 和 $B$ 的元素的新的集合。并集以符号 $A \cup B$ 表示，读作“$A$ 并 $B$”。

**例1** 若 $A=\{3,5,7\}$ 且 $B=\{2,3,4,8\}$ 则

$$A \cup B=\{2,3,4,5,7,8\}$$

此例描述了这样一种情况：两个集合既不相等，也非不相交，且互不为对方子集。

**例2** 再看图 2.1，我们知道整数集合与分数集合的并集是有理数集合；同样，有理数集合与无理数集合的并集构成了所有实数集合。

另外，两个集合 $A$ 和 $B$ 的交是一个新的集合，它包含且仅包含同时属于 $A$ 和 $B$ 的那些元素。交集以符号 $A \cap B$ 表示，读作“$A$ 交 $B$”。

**例3** 对于例 1 中的集合 $A$ 和 $B$，我们可以写成

$$A \cap B=\{3\}$$

**例4** 若 $A=\{-3,6,10\}$，$B=\{9,2,7,4\}$，则 $A \cap B=\varnothing$，即集合 $A$ 与 $B$ 不相交，因此其交集为空集——$A$ 和 $B$ 没有任何公共元素。

很明显，交集是一个比并集受到更多约束的概念，只有共同属于 $A$ 和 $B$ 的元素，才被前者所接受；然而无论属于 $A$ 或者 $B$ 的元

素，都可构成并集的元素。运算符号$\cap$和$\cup$——顺便说一句，它们同符号$\sqrt{\ }$、+、÷等具有同样的地位——分别具有“和”与“或”的含义。比较下面两个交集和并集的正式定义，可更好地理解这一点：

交集

$$A \cap B = \{x \mid x \in A \text{ 和 } x \in B\}$$

并集

$$A \cup B = \{x \mid x \in A \text{ 或 } x \in B\}$$

在解释补集的概念以前，我们首先引入全集的概念，在所讨论的特定范围内，如果我们仅把前7位正整数作为一个集合，并将其视为全集$U$，那么，对于一个给定集合，比如$A=\{3,6,7\}$，我们可以定义另一个集合$\tilde{A}$，它由包含在全集$U$但不包含在集合$A$中的所有元素构成，读作“$A$的补集”。即， 15

$$\tilde{A} = \{x \mid x \in U \text{ 和 } x \notin A\} = \{1,2,4,5\}$$

注意，符号$\cup$有“或”的含义，符号$\cap$意味着“和”，而补集符号～则有“不”的含义。

**例5**　若$U=\{5,6,7,8,9\}$和$A=\{5,6\}$，那么$\tilde{A}=\{7,8,9\}$。

**例6**　何为全集$U$的补集？因为所考察的每一对象（数）均在全集之中，所以$U$的补集必定为空集。即$U=\emptyset$。

集合的三种运算可以图2.2更直观地表示，此图被称为维恩图。图$a$中，上面圆中的点组成集合$A$，下面圆中的点组成集合$B$，而集合$A$和$B$的并集则由覆盖两个圆的阴影部分所组成。在图$b$中，给出同样两个集合（两个圆）。因为两集合的交仅包含共属于两集合的部分，所以只有两个集合的交叉部分（阴影）符合定义要求。在图$c$中，令矩形内的所有点为全集，圆内的所有点为集合$A$，则补集$\tilde{A}$则是圆以外的阴影区域。

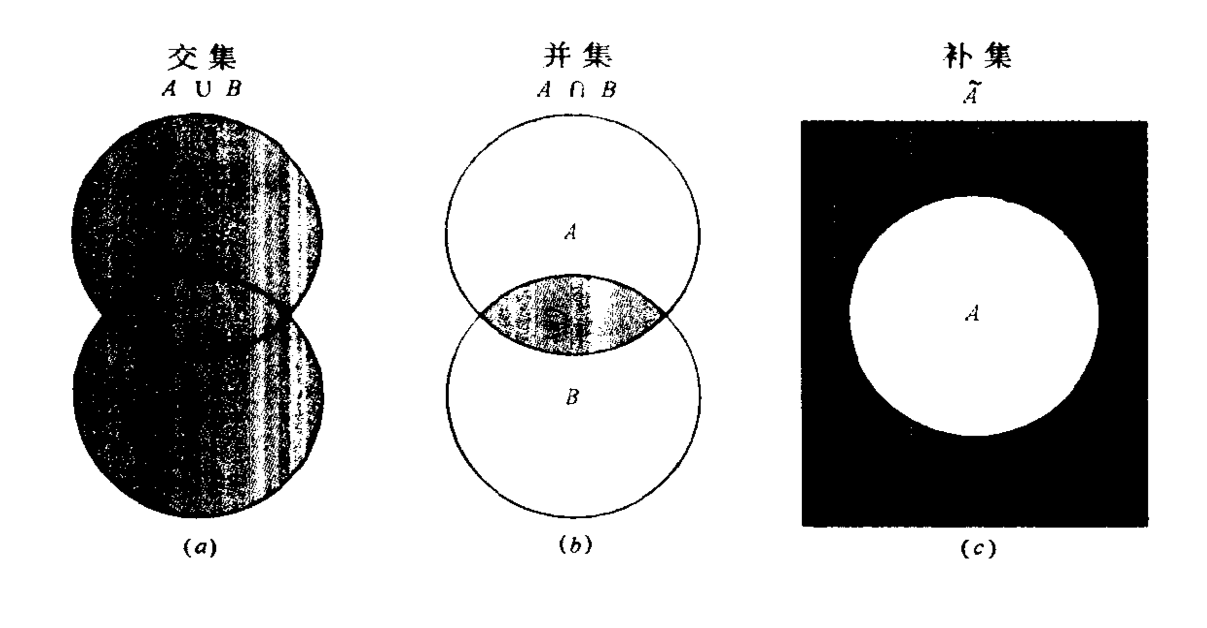



图 2.2

## 集合的运算法则

需要指出的是,从图 2.2 可知,图形 $a$ 的阴影面积不仅是 $A\cup B$,而且也是 $B\cup A$。类似地,图形 $b$ 中的小阴影面积不仅直观地表示 $A\cap B$,而且同样表示 $B\cap A$,将此结果规律化,可称之为并集和交集的**交换律**

16

$$A \cup B = B \cup A \quad A \cap B = B \cap A$$

这些关系非常类似于代数运算法则 $a+b=b+a$ 及 $a\times b=b\times a$。

要取三个集合 $A$、$B$、$C$ 的并集,可以先取任意两个集合的并集,再将其与第三个集合取并集。类似的过程也可用于交集的运算。这些运算结果在图 2.3 中加以说明。有趣的是,运算过程中集合选择的次序是无关紧要的,这一事实便引出了并集与交集的**结合律**:

$$A \cup (B \cup C) = (A \cup B) \cup C$$

$$A \cap (B \cap C) = (A \cap B) \cap C$$

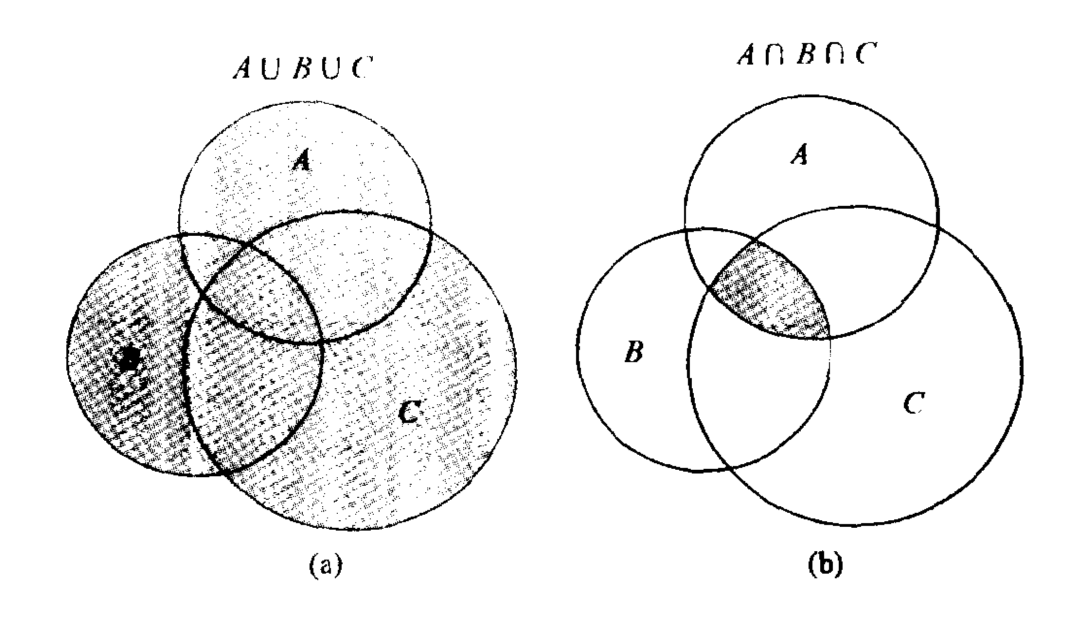



图 2.3

这些方程极易使我们想起代数运算法则 $a+(b+c)=(a+b)+c$，以及 $a\times(b\times c)=(a\times b)\times c$。

在联合运用并集和交集时，还要应用另一条法则，即并集和交集的分配律：

$$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$$
$$A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$$

它们类似于代数运算法则 $a\times(b+c)=(a\times b)+(a\times c)$。

**例 7** 已知 $A=\{4,5\}$，$B=\{3,6,7\}$，$C=\{2,3\}$，验证分配律，为验证此规律的前一部分，我们分别求出左边和右边的表达式

左： $A\cup(B\cap C)=\{4,5\}\cup\{3\}=\{3,4,5\}$

右： $(A\cup B)\cap(A\cup C)=\{3,4,5,6,7\}\cap\{2,3,4,5\}=\{3,4,5\}$

因两边结果相同，所以第一部分得证，重复此过程，可验证规律的第二部分，我们有

左 $A\cap(B\cup C)=\{4,5\}\cap\{2,3,6,7\}=\emptyset$ 17

右 $(A\cap B)\cup(A\cap C)=\emptyset\cup\emptyset=\emptyset$

所以分配律得证。

## 练习 2.3

1 用集合符号写出下列集合：

(a) 大于 27 的所有实数集；

(b) 大于 8 但小于 73 的所有实数集。

2 给定集合 $S_1=\{2,4,6\}$，$S_2=\{7,2,6\}$，$S_3=\{4,2,6\}$，$S_4=\{2,4\}$，下面哪些说法正确？

(a) $S_1=S_2$　　(d) $3\notin S_2$　　(g) $S_1\supset S_4$

(b) $S_1=R$　　(e) $4\notin S_4$　　(h) $\emptyset\subset S_2$

(c) $5\in S_2$　　(f) $S_4\subset R$　　(i) $S_3\supset\{1,2\}$

3 根据上题给出的四个集合，求

(a) $S_1\cup S_2$　　(c) $S_2\cap S_3$　　(e) $S_4\cap S_2\cap S_1$

(b) $S_1\cup S_3$　　(d) $S_2\cap S_4$　　(f) $S_3\cup S_1\cup S_4$

4 下述哪些说法是正确的？

(a) $A\cup A=A$　　(e) $A\cap\emptyset=\emptyset$

(b) $A\cap A=A$　　(f) $A\cap U=A$

(c) $A\cup\emptyset=A$　　(g) $\bar{A}$ 的补集是 $A$

(d) $A\cup U=U$

5 已知集合 $A=\{4,5,6\}$，$B=\{3,4,6,7\}$，$C=\{2,3,6\}$，验证分配律。

6 用维恩图法，根据逐次形成阴影的不同顺序，验证分配律。

7 列举集合 $\{a,b,c\}$ 的所有子集。

8 列举集合 $S=\{1,3,5,7\}$ 的所有子集。共有多少个？

9 例 6 表明 $\emptyset$ 是 $U$ 的补集，但因空集是任意集合的子集，所以它必定是 $U$ 的子集。因为“$U$ 的补集”这个术语具有“不在 $U$ 中”的含义，而“$U$ 的子集”这一术语又有“包含在 $U$ 中”的含义。$\emptyset$ 同时具有这两种含义，似乎是自相矛盾的，你如何解释这个谜？

## 2.4 关系与函数

通过使用数系中各类数的概念推进了我们对集合的讨论。但集合的元素不仅可以是数,也可以是其它对象。特别地,我们可以运用“有序偶”集合这一术语(下面将定义),引出关系和函数的重要概念。 18

### 有序偶

根据定义$\{a,b\}=\{b,a\}$,所以在写出集合$\{a,b\}$时,我们并不关注元素$a$、$b$出现的顺序。在此情况下,这对元素$a$和$b$是**无序偶**。但是当$a$和$b$的顺序具有重要意义时,我们可以写出两对不同的有序偶,以$(a,b)$和$(b,a)$表示,且具有$(a,b)\neq(b,a)$的性质,$a=b$时除外,类似的概念还可以应用于两个以上元素的集合;在此情况下我们可以区分有序三元组和无序三元组,有序四元组和无序四元组,有序五元组和无序五元组等。有序偶,有序三元组等统称**有序集合**。

**例 1** 为表示某班学生的年龄和体重,我们可以构造一个有序偶$(a,w)$,其中第一个元素表示学生年龄(以岁为单位),第二个元素表示体重(以磅为单位)。则(19, 127)和(127,19)具有明显不同的含义。且后一有序偶几乎不可能适合于任何一个学生。

**例 2** 当我们谈及一场竞赛的五个决赛选手集合时,其顺序是无关紧要的,所以我们有一个无序五元组,但当他们经过决赛,排出冠军、亚军等名次后,这种排列则变成了一个有序五元组。

同其它对象一样,有序组可以作为集合的元素。考察图 2.4 中

的直角坐标即笛卡尔坐标平面，$x$ 轴和 $y$ 轴垂直相交，将一平面分成四个象限。这个 $xy$ 平面是一个无限点集，每一个点代表一个有序偶，其中第一个元素代表 $x$ 的值，第二个元素代表 $y$ 的值。显然，标以(4,2)的点不同于标以(2,4)的点，在这里顺序是重要的。

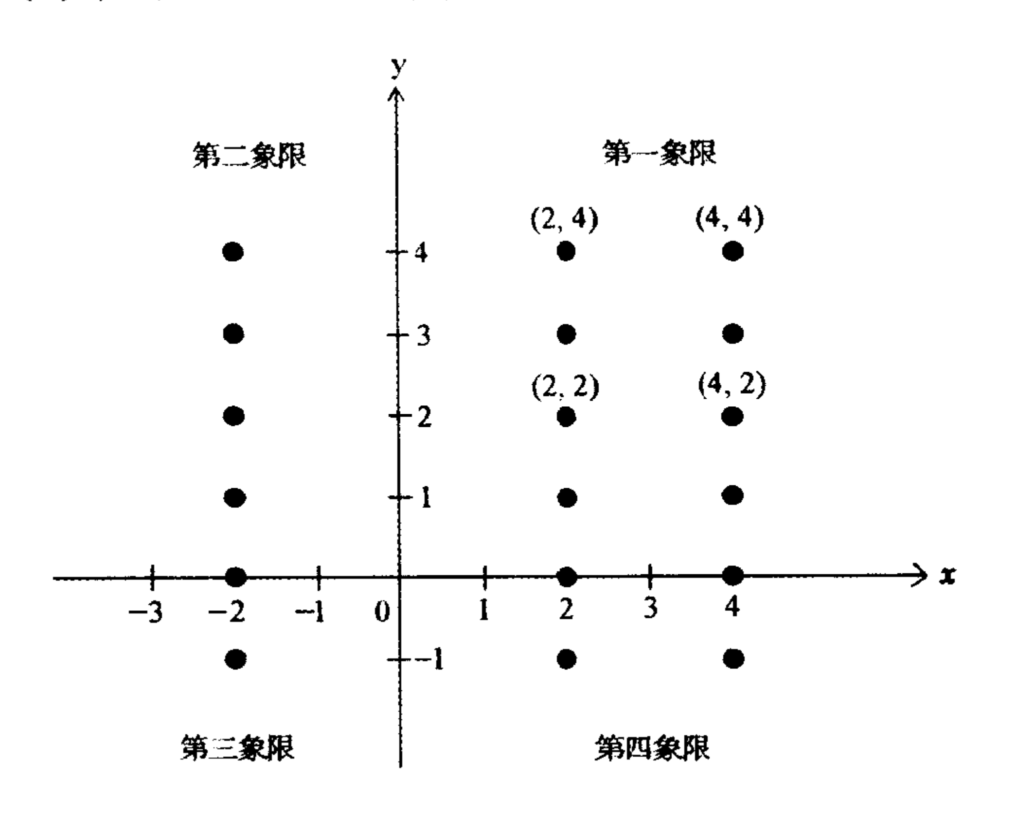



图 2.4

根据这种直观的理解，我们准备考察有序偶的生成过程。假设由两个已知的集合 $x = \{1,2\}$，$y = \{3,4\}$，我们希望形成所有的有序偶集合，其中第一个元素取自集合 $x$，第二个元素取自集合 $y$。这样我们便得到 4 个有序偶构成的集合(1,3)，(1,4)，(2,3)和(2,4)。此集合被称作 $x$ 和 $y$ 的笛卡尔积(以笛卡尔名字命名)，或称为 $x$ 和 $y$ 的直积，以 $x \times y$ 表示，读作"$x$ 叉乘 $y$"。需要重点记忆的是，当 $x$ 和 $y$ 是数集时，笛卡尔积便是有序偶。用列举法或描述法，我们可分别将笛卡尔积表示成

$$x \times y = \{(1,3),(1,4),(2,3),(2,4)\}$$

或 $$x \times y = \{(a,b) \mid a \in x \text{ 和 } b \in y\}$$

后者实际上可作为任意给定集合 $x$ 和 $y$ 的笛卡尔积的一般定义。[19]

为开拓视野，现在我们令 $x$ 和 $y$ 包括所有的实数，则相应的笛卡尔积为

$$x \times y = \{(a,b) \mid a \in R \text{ 和 } b \in R\} \tag{2.3}$$

将表示具有实值元素的全部有序偶的集合。此外，每个有序偶对应图 2.4 笛卡尔坐标平面上的**唯一**的点。反之，笛卡尔坐标平面上的每一个点也对应于集合 $x \times y$ 中的**唯一**一个有序偶。根据这两种唯一性，我们说笛卡尔积(2.3)中的有序偶集与平面直角坐标中的点集之间存在**一一对应性**。我们现在易于理解 $x \times y$ 这一概念了，因为我们可以将其与图 2.4 中 $x$ 轴与 $y$ 轴的交叉联系起来。表示(2.3)中的集合($x \times y$)的更为简洁的方式是将其直接写成 $R \times R$，它通常也表示成 $R^2$。

拓展这一思想，我们也可以将三个集合 $x$，$y$ 和 $z$ 的笛卡尔积定义为

$$x \times y \times z = \{(a,b,c) \mid a \in x,\ b \in y, c \in z\}$$

这是一个有序三元组的集合。进而，如果集合 $x$，$y$ 和 $z$ 均由全部实数构成，则笛卡尔积将对应于三维空间的所有点的集合，这可以用 $R \times R \times R$ 表示，或更简单地表示成 $R^3$。在后面的推广中，所[20]有变量均取实值，因而我们的一般性讨论均表示成 $R^2$，或 $R^3$，…，或 $R^n$。

## 关系与函数

因为任意有序偶将 $x$ 值和 $y$ 值联系起来，任意有序偶的集

聚——笛卡尔积(2.3)的任意子集——将包含 $x$ 和 $y$ 的关系。给定一个 $x$ 值,将根据这种关系确定另一个或多个 $y$ 值。为方便起见,我们将 $x \times y$ 的元素一般写成$(x,y)$。不像(2.3)中那样写成$(a,b)$,这里的 $x$ 和 $y$ 均为变量。

**例 3** 集合$\{(x,y) \mid y=2x\}$是一个有序偶集合,比如它包括$(1,2)$,$(0,0)$和$(-1,-2)$。这个集合包含一种关系,其对应的图形部分是图 2.5 中的直线 $y=2x$ 上的点的集合。

**例 4** 集合$\{(x,y) \mid y \leqslant x\}$,包含如$(1,0)$,$(1,1)$,$(1,-4)$这样的有序偶,它构成了另一种关系。在图 2.5 中,这个集合对应于满足不等式 $y \leqslant x$ 的阴影面积的所有点的集合。

21 当 $x$ 值给定时,未必一定总能从某种关系中确定唯一的 $y$ 值。在例 4 中,三个典型的有序偶表明,若 $x=1$,则 $y$ 可以取 $n$ 个不同的值,如 0,1,或 $-4$,且满足给定的关系。从图形上看,满足这一关系的两个或多个点落在 $xy$ 平面中的一个垂线上。图 2.5 可以说明这一点。图中阴影区域(表示关系 $y \leqslant x$)中的许多点落在标有 $x=a$ 的一条垂直虚线上。

然而,作为一个特例,一种关系可以使得给定的 $x$ 值,仅存在一个对应的 $y$ 值,例 3 就是一例。在这种情况下,称 $y$ 为 $x$ 的一个函数,并以 $y=f(x)$表示,读作"$y$ 是 $x$ 的函数 $f$"。[注意 $f(x)$ 并不意味着 $f$ 乘以 $x$]因此,一个函数是具有这种性质的有序偶集合:$x$ 值惟一地确定一个 $y$ 值[①]。显然,一个函数一定是一种关系,但一种关系未必是一个函数。

---

① "函数"这一定义对应于我们以前所称的单值函数这一术语,这里所指的关系,以前正式的名称是多值函数。

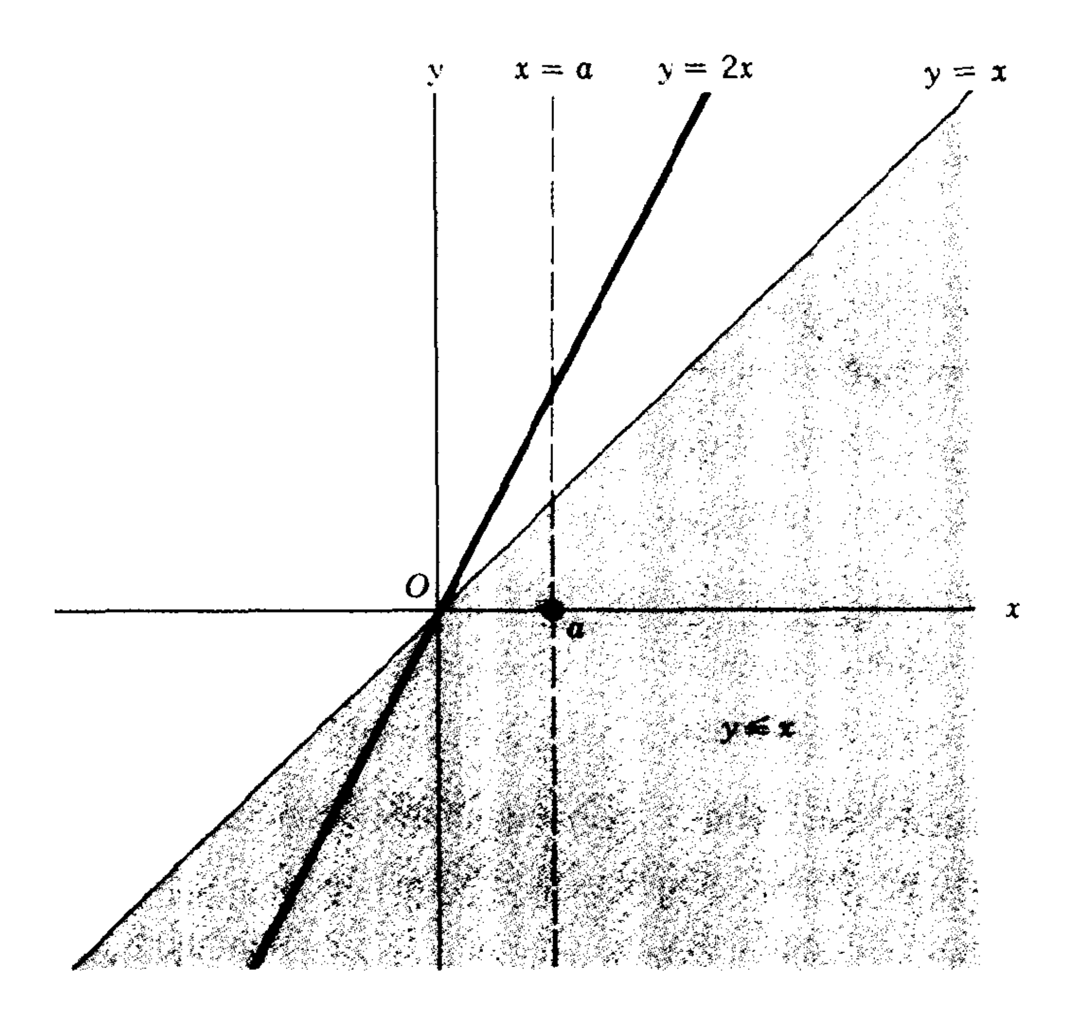



图 2.5

尽管函数的定义规定对于每一个 $x$ 有一个唯一的 $y$，但并不要求反之也成立，换句话说，多个 $x$ 值可与同一个 $y$ 值相联系。图 2.6 描述了这种可能性：图中根据函数 $y=f(x)$，集合 $x$ 中的 $x_1$ 和 $x_2$ 均与集合 $y$ 中的同一值($y_0$)相联系。

函数也称映射或变换，二者均意味着将一对象与另一对象联系起来的行动。表达式 $y=f(x)$ 中的函数符号 $f$ 则可以解释成一种规则：依此规则集合 $x$ 被“映射”(或变换)到集合 $y$。因此我们可以写成：

$$f: x \rightarrow y$$

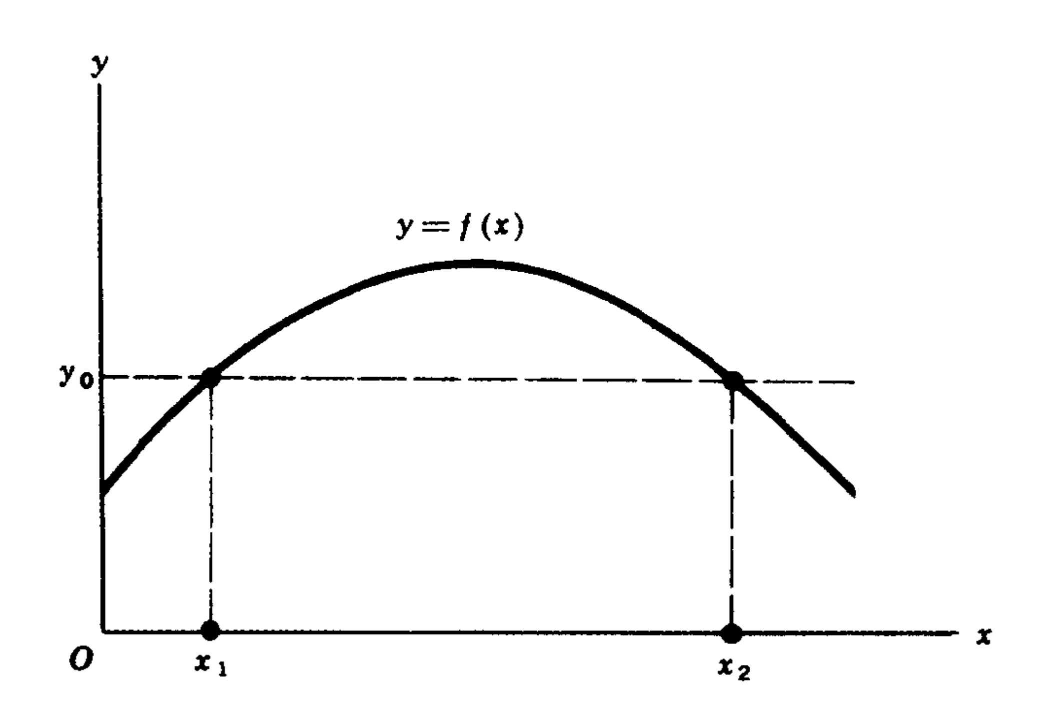



图 2.6

22其中箭头表示映射，字母 $f$ 表示映射遵循的规则。因为 $f$ 表示映射的特殊规则，所以必须用不同的函数符号去表示出现在同一模型中的另外一个函数。习惯使用的符号还有 $g$, $F$, $G$，希腊字母 $\phi$(phi)和 $\varphi$(psi)，以及它们的大写 $\Phi$ 和 $\Psi$。例如，两个变量 $y$ 和 $z$ 可能都是 $x$ 的函数，但若一个函数写成 $y=f(x)$，另一个则应写成 $z=g(x)$或者 $z=\phi(x)$。但写成 $y=y(x)$及 $z=z(x)$也是允许的，这样就省略了符号 $f$ 和 $g$。

在函数 $y=f(x)$中，$x$ 被称作函数的自变量，$y$ 则被称作函数值。我们还可以把 $x$ 称作自变量，把 $y$ 称作因变量。在给定的范围内，$x$ 能够取的所有值的集合称作函数的定义域，它可能是所有实数集合的一个子集。$x$ 所映射的 $y$ 值称作 $x$ 值的象。所有象的集合称作函数的值域，它是变量 $y$ 可以取的所有值的集合。所以定义域与自变量 $x$ 相联系，值域则与因变量 $y$ 相联系。

如图 2.7$a$ 所示，我们可以将函数 $f$ 视为一个映射规则，它将

某线段(定义域)上的每一点映射到另一线段(值域)上的某点。如图形 $b$ 所示,将定义域置于 $x$ 轴上,值域置于 $y$ 轴上,我们马上得到熟悉的二维图形,在此图形中,$x$ 值与 $y$ 值的关系是由诸如$(x_1, y_1)$和$(x_2, y_2)$等有序偶集合规定的。

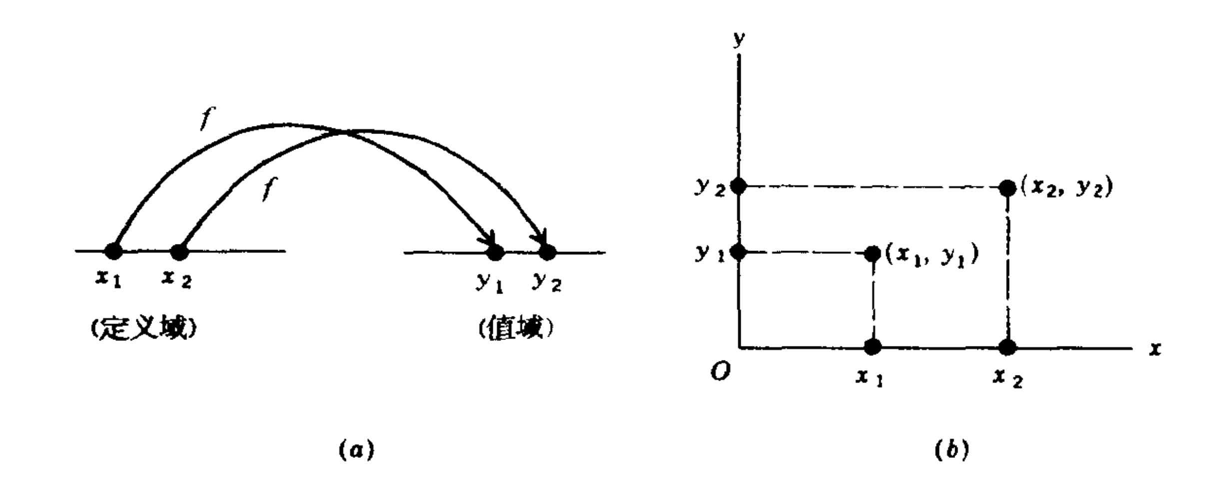



图 2.7

在经济模型中,行为方程通常以函数形式引入。由于经济模型中的大多数变量因其性质所限,一般取非负实数,[①] 所以其定义域也受到同样限制。这就是为什么大多数经济模型的几何表示仅画第一象限的原因。一般而言,我们不需为每一经济模型的每一函数设定定义域而伤脑筋。当未特别设定时,我们将定义域和值域理解为仅包括使函数具有经济意义的那些数值。

**例 5** 厂商每日的总成本 $C$ 为其日产出 $Q$ 的函数:$C = 150 + 7Q$。该厂商日最大产出为 100 单位。那么,成本函数的定义域和值域是什么? 因为 $Q$ 可在 0 与 100 之间变化,所以定义域为集合 $0 \leqslant Q \leqslant 100$,或更正式地 23

$$\text{定义域} = \{Q \mid 0 \leqslant Q \leqslant 100\}$$

① 如果可以取零值,我们就说"非负"而不说"正"。

至于值域，因函数的图形为一条直线，$C$ 的最小值为 150（当 $Q=0$），最大值为 850（当 $Q=100$），所以我们有

值域 $=\{C \mid 150 \leqslant C \leqslant 850\}$

但要注意，值域的最大值未必总是在定义域的最大值时达到。

**练习 2.4**

---

1 给定 $S_1=\{3, 6, 9\}$，$S_2=\{a, b\}$，及 $S_3=\{m, n\}$，求笛卡尔积：

(a) $S_1 \times S_2$，(b) $S_2 \times S_3$，(c) $S_3 \times S_1$

2 根据上题给出的信息，求笛卡尔积 $S_1 \times S_2 \times S_3$。

3 一般而言，$S_1 \times S_2 = S_2 \times S_1$ 是否成立？在何条件下，这两个笛卡尔积相等？

4 在直角坐标系内画出下列图形，它们是否表示一个函数？

(a) 圆　　(b) 三角形　　(c) 矩形

5 若函数 $y=5+3x$ 的定义域为集合 $\{x \mid 1 \leqslant x \leqslant 4\}$，求此函数的值域并以集合方式将其表示出来。

6 对于函数 $y=-x^2$，如果其定义域是所有非负实数的集合，那么其值域是什么？

---

## 2.5 函数的类型

表达式 $y=f(x)$ 是一个一般描述，它表明存在映射的可能，但并不明确映射的实际规则。下面我们考察几种特定的函数类型，每种类型均代表不同的映射规则。

## 常值函数

值域仅有一个元素的函数,称作常值函数。我们拿出一个函 24
数作为例子

$$y = f(x) = 7$$

也可将其表示成 $y=7$ 或 $f(x)=7$,无论 $x$ 取何值,其值始终为 7。在坐标平面上,这样的函数表现为一条水平线。在国民收入模型中,当投资($I$)为外生决定的,我们可以有下述形式的投资函数:$I=1$亿美元,或 $I=I_0$,它是常值函数的例子。

## 多项式函数

常值函数实际上是我们所说的多项式函数的一种“退化”。具有一个变量 $x$ 的多项式的一般形式为

$$y = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \tag{2.4}$$

其中每一项均包含一个系数和变量 $x$ 的一个非负整数幂。(本节后面将说明,一般我们可以写成 $x^1=x$, $x^0=1$,所以上式前两项可以分别写成 $a_0x^0$ 和 $a_1x^1$。)注意,我们使用标有下标的符号 $a_0$, $a_1$, $a_2$,⋯ $a_n$,而不用符号 $a$, $b$, $c$⋯来表示系数。这基于两方面的考虑:(1) 我们可以节约符号,因为以此方式表示,仅有一个符号 $a$ 就够用了;(2) 下标可以帮助我们确定一个特定系数在整个方程中的位置。如在(2.4)式中,$a_2$ 是 $x^2$ 的系数,等等。

依赖于整数 $n$ 的值(它设定了 $x$ 的最高幂数),我们可以有以下几种多项式函数的子类别:

当 $n=0$: $y=a_0$ [常值函数]

当 $n=1$：$y=a_0+a_1x$　　　　　　　　　[线性函数]

当 $n=2$：$y=a_0+a_1x+a_2x^2$　　　　　　[二次函数]

当 $n=3$：$y=a_0+a_1x+a_2x^2+a_3x^3$　[三次函数]

等等。$x$ 的幂次的上标值称为指数。函数的最高幂次，即 $n$ 值，被称作多项式函数的次数；比如，二次函数称作二次多项式，三次函数称作三次多项式等。[1] 等号右边的各项的排列顺序是无关紧要的，也可以将其降幂排列。尽管方程左边我们运用符号 $y$，但也可以用 $f(x)$ 来代替。

如图 2.8$a$ 所示，在直角坐标平面内，线性函数是一条直线。当 $x=0$ 时，线性函数得到 $y=a_0$，有序偶 $(0, a_0)$ 位于此直线上，这样我们得到所谓的"$y$ 截距"（或纵截距），因为它是纵轴与直线的交点。另一个系数 $a_1$，测度直线的斜率（倾斜的程度）。这意味着 $x$ 增加一单位，将会导致 $y$ 增加 $a_1$ 单位。图 2.8$a$ 所描述的是 $a>0$ 的情况，斜率为正，因此是一条向上倾斜的直线；若 $a<0$，则直线将向下方倾斜。

另一方面，二次函数的图形为抛物线——或粗略地讲，是一条内部只有一个折返或弯曲的曲线。图 2.8$b$ 所描述的特定情况中，$a_2$ 为负；当 $a_2>0$ 时，曲线向另一方向张开，状似山谷，而非山峰。三次函数的图形如图 2.8$c$ 所示，一般会出现两次弯曲。这些函数在下面讨论的经济模型中要经常使用。

---

① 刚才提到的几个方程，最后一项的系数 $a_n$ 总是假设不为零，否则此函数将退化为低次多项式。



线性函数
$y = a_0 + a_1 x$

斜率 $= a_1$

$a_0$

$O$ $x$ $y$

(a)

二次函数
$y = a_0 + a_1 x + a_2 x^2$

($a_2 < 0$ 的情况)

$a_0$

$O$ $x$ $y$

(b)

三次函数
$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$

$a_0$

$O$ $x$ $y$

(c)

等轴双曲线
$y = \frac{a}{x}$

($a > 0$)

$O$ $x$ $y$

(d)

指数函数
$y = b^x$

($b > 1$)

$O$ $x$ $y$

(e)

对数函数
$y = \log_b x$

$O$ $x$ $y$

(f)

图 2.8

## 有理函数

下面这样一个函数

$$y = \frac{x - 1}{x^2 + 2x + 4}$$

$y$ 被表示成变量 $x$ 的两个多项式的比率,称作有理函数(同样,"有理"仅意味着比率),根据这一定义,任何多项式函数本身必定为有理函数,因为它总可以表示成该式与常值函数 1 之比。

在经济学中有重要应用的一个特殊有理函数是

$$y = \frac{a}{x} \text{或} xy = a$$

如图 2.8$d$ 所示,绘出的图形是等轴双曲线。因为在此情况下,两个变量之积总为一固定常数,此函数可以用于表示一种特定的需求曲线,——价格 $P$ 和需求量 $Q$ 作为两轴——在所有价格水平上,总支出不变。(这种需求曲线是一种在曲线的每一点都有单一需求弹性的曲线。)另一个应用是平均固定成本曲线($AFC$)。令 $AFC$ 为坐标系的一轴,产出 $Q$ 为另一轴,$AFC$ 曲线一定是条等轴双曲线,因为 $AFC \times Q$(= 总固定成本)是一个固定常数。

等轴双曲线 $xy = a$ 即使无限向上和向右延伸,也不会与两轴相交。或者说,曲线将逐渐趋近于两轴:随着 $y$ 的增大,曲线将越来越接近 $y$ 轴,但实际上永远不可能与之相交;对于 $x$ 轴来说,也是同理。这两轴构成函数的渐近线。

27

## 非代数函数

任何以多项式和(或)多项式的根(如平方根)表示的函数,均

是代数函数。因此，我们到目前为止所讨论的函数均为代数函数。诸如 $y=\sqrt{x^2+3}$ 这样的函数并非有理函数，但是代数函数。

然而，像 $y=b^x$ 这样的指数函数，其自变量出现在指数上，称非代数函数。与其密切相关的对数函数，如 $y=\log_b x$，也是非代数函数。这两类函数将在第 10 章中详细讨论，但在图 2.8$e$ 和图 2.8$f$ 中我们绘出其一般图形。其它类型的非代数函数有三角函数，我们将在第 15 章与动态分析相关的内容中讨论。我们还要加上一句，更为深奥的超越代数也是非代数函数。

## 关于指数的额外解释

在讨论多项式函数时，我们引入了指数这一术语，用以表示一个变量（或者数）的自乘幂数。表达式 $6^2$ 表示 6 自乘到二次幂，即 6 被其自己相乘，或 $6^2 \equiv 6\times 6=36$。一般而言，我们定义

$$x^n \equiv \underbrace{x\times x\times\cdots\times x}_{n\text{ 项}}$$

且作为特殊情况，我们要注意 $x^1=x$。由此一般定义，指数服从以下法则：

**法则 I** $x^m\times x^n=x^{m+n}$（例如 $x^3\times x^4=x^7$）

证明 $$x^m\times x^n=(\underbrace{x\times x\times\cdots\times x}_{m\text{项}})(\underbrace{x\times x\times\cdots\times x}_{n\text{项}})=\underbrace{x\times x\times\cdots\times x}_{m+n\text{项}}=x^{m+n}$$

**法则 II** $\dfrac{x^m}{x^n}=x^{m-n}\ (x\neq 0)$ （例如 $\dfrac{x^4}{x^3}=x$）

28

证明 $$\frac{x^m}{x^n} = \frac{\overbrace{x \times x \times \cdots \times x}^{m\text{项}}}{\underbrace{x \times x \times \cdots \times x}_{n\text{项}}} = \underbrace{x \times x \times \cdots \times x}_{m-n\text{项}} = x^{m-n}$$

因为分母中的 $n$ 项消去了分子中的 $m$ 项。注意当 $x=0$ 时,该法则不成立。因为当 $x=0$ 时,表达式 $x^m/x^n$ 以零作除数,没有意义。

如果 $m<n$,比如 $m=2, n=5$,会如何?在此情况下,依法则 II,$x^{m-n}=x^{-3}$,$x$ 的幂次为负。这意味着什么呢?答案仍由法则 II 自己给出:当 $m=2, n=5$ 时,我们有

$$\frac{x^2}{x^5} = \frac{x \times x}{x \times x \times x \times x \times x} = \frac{1}{x \times x \times x} = \frac{1}{x^3}$$

因此 $x^{-3}=1/x^3$。可以将其推广为另一法则:

**法则 III** $$x^{-n} = \frac{1}{x^n} \quad (x \neq 0)$$

一个非零数的负 $n$ 次幂等于该数 $n$ 次幂的倒数。

应用法则 II 时的另一种特殊情况是 $m=n$,它使得表达式 $x^{m-n}=x^{m-m}=x^0$。为解释 $x$ 的零次幂的含义,我们可以按法则 II 写出 $x^{m-m}$ 项,会得到 $x^m/x^m=1$。因此我们可以得出这一结论:任何非零数的零次幂等于 1。(表达式 $0^0$ 无意义。)可将此表示成另一法则

**法则 IV** $$x^0 = 1 \quad (x \neq 0)$$

只要我们仅涉及多项式函数,$x$ 的(非负)整数次幂已足以满足需要。但在指数函数中,指数是一个也可取非整数的变量。为解释像 $x^{1/2}$ 这样的数的含义,我们考察这样一个事实:由法则 I,我们有

$$x^{1/2} \times x^{1/2} = x^1 = x$$

因为 $x^{1/2}$ 被自乘等于 $x$，$x^{1/2}$ 一定是 $x$ 的平方根，类似地，$x^{1/3}$ 一定是 $x$ 的立方根。因此，一般而言，我们可表述下列法则：

**法则 V** $x^{1/n} = \sqrt[n]{x}$

指数服从的另两个法则是

**法则 VI** $(x^m)^n = x^{mn}$

**法则 VII** $x^m \times y^m = (xy)^m$

29

## 练习 2.5

---

1 用图形表示下列函数

(*a*) $y = 8 + 3x$ (*b*) $y = 8 - 3x$ (*c*) $y = 3x + 12$

(每题仅考察定义域为非负实数的情况。)

2 上面(*a*)题与(*b*)题的主要区别是什么？这种区别在图形上如何反映出来？(*a*)与(*c*)的主要区别是什么？这种区别在图形上如何反映出来？

3 用图形表示下列函数。其定义域为 $-5 \leqslant x \leqslant 5$ 的值的集合。

(*a*) $y = -x^2 + 5x - 2$ (*b*) $y = x^2 + 5x - 2$

我们已经知道 $x^2$ 的系数的符号决定二次函数的图形状似"山峰"或似"山谷"。就现在这两个问题而言，哪一个符号与山峰有联系？给出直观解释。

4 假设 $x$ 和 $y$ 仅取正值，用图形表示函数 $y = 36/x$，然后再假设两变量也可取负值，图形要作何调整以反映假设的变化？

5 简化下列各式：

(*a*) $x^4 \times x^{15}$ (*b*) $x^a \times x^b \times x^c$ (*c*) $x^3 \times y^3 \times z^3$

6 求：(*a*) $x^3/x^{-3}$ (*b*) $(x^{1/2} \times x^{1/3})/x^{2/3}$

7 证明 $x^{m/n} = (\sqrt[n]{x})^m$。指明在每一步骤所应用的法则。

8 证明法则 VI 和法则 VII。

---

## 2.6 两个或两个以上自变量的函数

迄今为止，我们仅考察了只有一个自变量的情况，$y=f(x)$。但函数的概念极易推广到具有两个及两个以上自变量的情况。给定函数：

$$z = g(x, y)$$

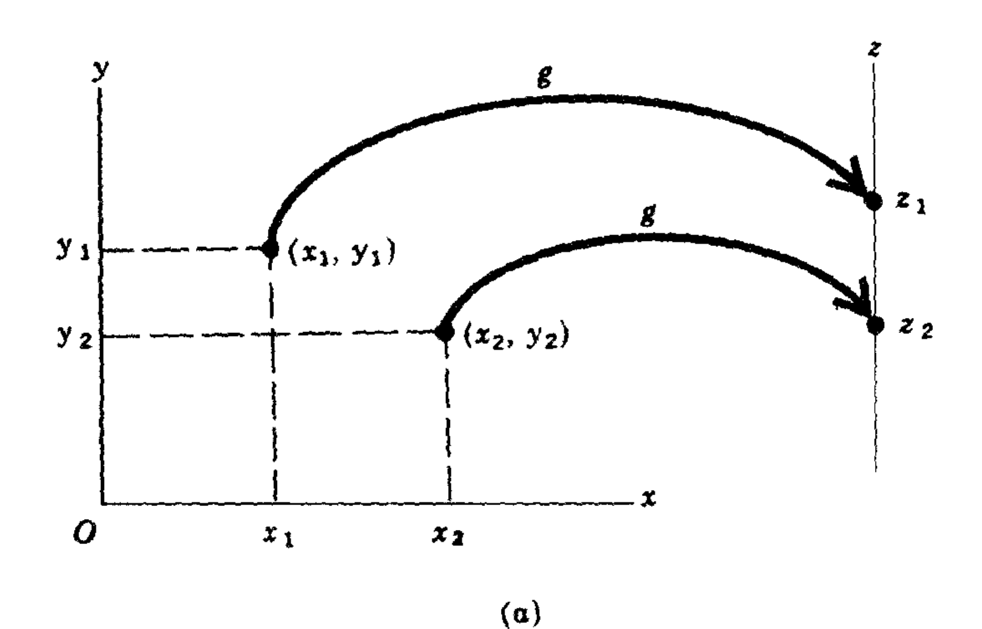
一对给定的 $x$ 值和 $y$ 值将惟一地确定一个因变量 $z$ 的值。这样的函数例子有

$$z = ax + by \text{ 或 } z = a_0 + a_1x + a_2x^2 + b_1y + b_2y^2$$

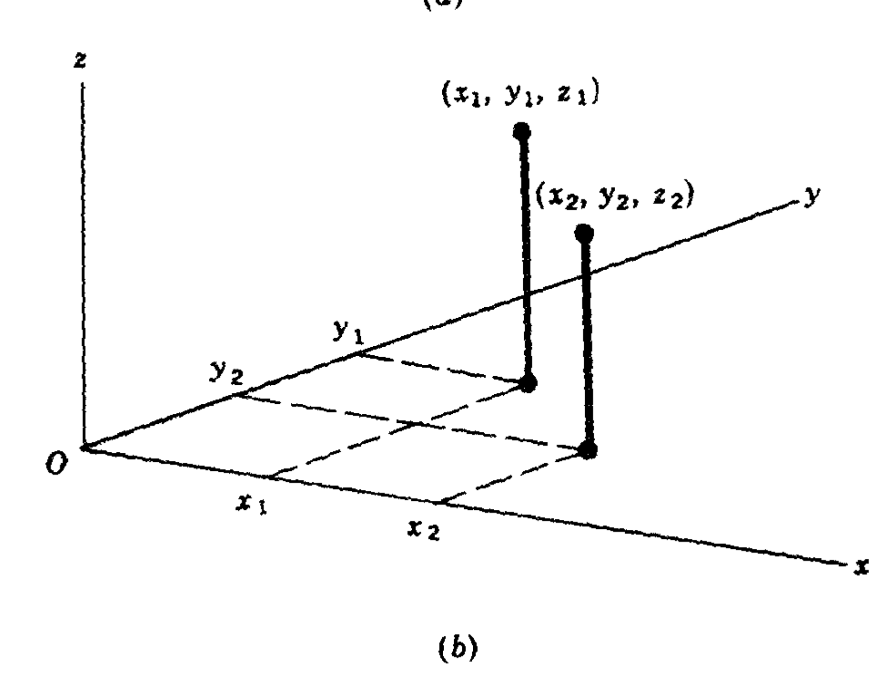
函数 $y=f(x)$ 把定义域中点映射到值域中的点，函数 $g$ 也同样如此。但是这里的定义域不再是一个数集，而是有序偶 $(x, y)$ 的集合，因为只有同时给定 $x$ 和 $y$ 值之后，才能确定 $z$ 值。因而函数 $g$ 是从二维空间中的点映射到线段上的点（即一维空间上的点）。如图 2.9$a$ 所示，从点 $(x_1, y_1)$ 映射到点 $z_1$，从点 $(x_2, y_2)$ 映射到 $z_2$ 等等。

如图 2.9$b$ 所示，如果作纵轴 $z$ 垂直于 $xy$ 平面，则产生一个三维空间。我们在这个三维空间中给出函数 $g$ 的如下几何解释：30 函数的定义域是 $xy$ 平面上的点的某个子集，定义域中给定点，比如 $(x_1, y_1)$ 的函数值（$z$ 的值），由该点垂直线段的高度表示。三个变量之间的联系可通过一个有序三元组 $(x_1, y_1, z_1)$ 来概括，它是三维空间中的一个特定的点。有序三元组的轨迹是一个曲面，构成了函数 $g$ 的图形。函数 $y=f(x)$ 是一个有序偶的集合，而函数 $z=g(x, y)$ 是一个有序三元组的集合。在经济模型中，这类函数我们将经常用到。一个现成的应用是在生产函数领域。假

设产量由资本($K$)和劳动($L$)的数量决定,则我们可以将生产函




(a)




(b)

图 2.9

数的一般形式写成:$Q = Q(K, L)$。 31

进一步将函数推广到三个或多个自变量的可能性是不言自明的。例如,对于函数 $y = h(u, v, w)$,我们可以将三维空间中的点

$(u_1, v_1, w_1)$映射到一维空间的点 $y_1$ 上,这样的函数可以用于表示消费者效用函数,即消费者的效用是其消费的三种不同商品的函数,其映射是从三维商品空间到一维效用空间。但此时不可能画出函数的图形,因为画出四维图形要画出有序四元组,但我们生活的这个世界仅是三维的。但无论如何,凭借直观的几何推理,我们仍可以将有序四元组$(u_1, v_1, w_1, y_1)$视为四维空间中的一个点。这些点的轨迹给出函数 $y = h(u, v, w)$的图形(不能画出),我们称其为超平面。这些术语,即点和超平面,也可推广到 $n$ 维空间的一般情况。

一个以上自变量的函数也可以分成不同类型。如下述形式的函数

$$y = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

是线性函数,其特点是每一个自变量均为一次幂。而二次函数则包含一个或多个自变量的一次幂和二次幂,但任何一项的变量的指数和不能超过 2。

注意,我们不用 $x, u, v, w$ 等表示自变量,而用符号 $x_1, x_2, \cdots, x_n$ 表示自变量。后一种符号,像系数的下标一样,具有节省符号,易于计算函数中包含的变量的数量等优点。

## 2.7 一般性水平

在讨论各种类型的函数时,我们没有明确关注所介绍函数例子的一般性水平变化的有关情况。在有些例子中,我们曾写出下述形式的函数:

$$y = 7 \quad y = 6x + 4 \quad y = x^2 - 3x + 1(等)$$

这些函数不仅都以数字系数表示，而且也都具体指明了每个函数是常值函数、线性函数还是二次函数。以图形看，每个这样的函数都给出了一明确定义的唯一的曲线。鉴于这些函数的数字性质，以此为基础的模型的解也将以数值形式发现。这些函数的缺点在于：如果我们想知道当出现不同的数值系数集合时我们的分析结论会如何变化，我们必须每次都重新进行我们的推理过程。因而，32 从特定函数得到的结果缺乏一般性。

在更一般水平的讨论和分析中，有下列形式的函数

$y=a \qquad y=a+bx \qquad y=a+bx+cx^2$ 等。

由于使用了参数，每个函数不是代表一条曲线，而是代表一族曲线。例如 $y=a$，不仅代表一些特定的情况，如 $y=0$，$y=1$ 及 $y=2$ 等，而且也表示 $y=\frac{1}{3}$，$y=-5$，…等无穷尽的情况。运用参数函数，数学运算的结果也可以参数表示。在下述意义上，这个结果更具有一般性：赋予模型解中的参数以不同的值，不必再重复推理过程，就可以得到所有具体答案。

为获得更高水平的一般性描述，我们可以运用一般函数形式 $y=f(x)$ 或 $z=g(x,y)$。当运用这种形式时，函数就不再局限于线性函数、二次函数、指数函数或者三角函数。所有这些函数均可纳入这个概念的框架内。因此，基于这种一般公式的分析结论最具一般性。然而，为了获得具有经济意义的结果，常需要对纳入模型的一般函数施加某些性质上的限制，比如限定需求函数的斜率应当为负，消费函数具有小于 1 的正的斜率等。

概括一下本章的内容，数学经济模型的结构已经清楚了。一般而言，它由方程组构成，这些方程可能是行为方程、定义方程或

者具有均衡条件性质的方程。[①] 行为方程通常以函数形式存在，函数可能是线性的或非线性的，可能是数值的或参数的，还可能有一个自变量或多个自变量等等。正是通过这些函数，模型所采纳的分析假设才得以给出数学表达。

因此，开始分析问题的第一步是为模型选择合适的内生变量和外生变量。第二步我们必须把所选定的关于环境中人类、组织、技术、法律以及其它有关方面的行为的分析假设转化为方程。这些环境影响着变量的运动。自此以后，我们便可以通过有关的数学运算和处理推导出一系列的结论，并给出合适的经济解释。

---

① 不等式也可能是模型的重要组成部分，但此时我们不予以考虑。

# 第 二 篇

# 静态(或均衡)分析

35

# 第 3 章　经济学中的均衡分析

我们将上一章介绍的分析程序首先应用于静态分析或均衡分析。为此，我们必须先明确地理解"均衡"的含义。

## 3.1　均衡的含义

像其它经济术语一样，均衡这一概念也可以不同方式来定义。有一个定义将均衡定义为"选定的一组具有内在联系的变量经过彼此调整，从而使这些变量所构成的模型不存在内在变化倾向"[①]的一种状态。在此定义中，有几个词需要特别加以注意。首先，"选定的"一词强调了这样一个事实：确实存在一些变量，由于分析者的选择而未被包含于模型之中。因此，这里所讨论的均衡仅与选定的特定变量集合有关。如果模型扩大，包含了额外的变量，适合于较小模型的均衡状态便不再适用于这种新的状态。

其次，"内在联系"一词意味着为了实现均衡状态，模型中的所有变量必须同时处于静止状态；而且，每一变量的静止状态必须与 36 所有其它变量的静止状态相一致。否则某一(些)变量将会变化，并引起其它变量的连锁反应，均衡便不再存在。

---

① 弗里茨·马克卢普："均衡与非均衡：错位的愈合和造作的政治"，载《经济杂志》，1958 年 3 月，第 9 页。(重印于马克卢普：《经济语义学文集》，Prentice-Hall 出版公司，Englewood Cliffs，N. J. 1963。)

第三,"内在的"一词意指在定义均衡时,所涉及的静止状态仅以模型内部力量的平衡为基础,而假定外部因素不变。从运算角度说,这意味着参数和外生变量被视为常数。当外部因素实际变化时,将会导致定义在新参数值基础上的新均衡;但在定义新均衡时,还要假设新参数值保持不变。

从本质上看,一个特定模型的均衡,是以缺乏变化趋势为特征的一种状态,正是因为如此,均衡分析(更确切地说,研究什么是均衡状态)被称之为静态学。均衡意味着缺乏变化趋势这一事实,易使人们得出这一结论:均衡是事物的一种理想的或合意的状态,因为只有理想状态才会缺乏变化动力。这个结论是缺乏依据的。尽管某些均衡代表了某种理想状态和可值得追求的东西——如从厂商角度来看的利润最大化状态——可另一些均衡却是不理想的需加以回避的状态,如非充分就业的国民收入均衡水平等。唯一合理的解释是:均衡一旦达到且外力不发生变化时,就有维持不变的倾向。

我们将称之为目标均衡的那种理想均衡,将在第四篇和第六篇中作为最优化问题加以讨论。本章的讨论仅限于非目标均衡,这种均衡并不是由于对特定目标的刻意追求,而是由于非个人的或超个人的经济力量相互作用与调节所致。例如,在给定供求条件下的市场均衡和给定消费与投资方式下的国民收入均衡,均属此例。

## 3.2 局部市场均衡——线性模型

在静态均衡模型中,标准的问题是求出满足模型均衡条件的

一组内生变量的值。这是因为我们一旦确定了这组值,实际上也就确定了均衡条件。下面我们描述所谓的“局部均衡市场模型”,即在一个孤立市场中的价格决定模型。

**模型的构建**

因为仅考察一种商品,所以模型中只需包括三个变量:商品的需求量($Q_d$),商品的供给量($Q_s$),及该商品的价格($P$)。商品的需求量可以用每周多少磅等来度量,价格可用美元等来度量。选择完变量后,下一步的任务就是要对市场运行作出若干假设,首先,我们必须设定市场均衡模型必不可少的市场均衡条件。标准的假设是:当且仅当超额需求为零($Q_d - Q_s = 0$),即当且仅当市场出清时,市场就实现均衡。但这马上提出一个问题:供求数量 $Q_d$ 和 $Q_s$ 本身是如何决定的?要回答此问题,我们需假设 $Q_d$ 是 $P$ 的递减线性函数(当 $P$ 增加时,$Q_d$ 减少);而 $Q_s$ 被假定为 $P$ 的递增线性函数($P$ 增加时,$Q_s$ 也随之增加),并满足这个条件:除非价格超过某一特定的正的价格水平,否则不会有商品供给。这样,模型将包括一个均衡条件,两个行为方程。这两个行为方程分别决定着市场供给和需求两个方面。

将其转化为数学表述,模型可以写成:

$$
\begin{aligned}
Q_d &= Q_s \\
Q_d &= a - bP \qquad (a, b > 0) \\
Q_s &= -c + dP \qquad (c, d > 0) \qquad (3.1)
\end{aligned}
$$

两个线性函数中出现的四个参数 $a$、$b$、$c$ 和 $d$,均设定为正。当画出需求函数曲线时,如图 3.1 所示,它与纵轴相交于 $a$,其斜率正如所要求的那样,为负,即 $-b$。供给函数也具有符合要求的斜

率，$d$ 值为正，但它与纵轴交于 $-c$，为什么我们要设定这样一个负的截距呢？因为只有这样做，我们才能使得供给曲线与横轴相交于正值 $P_1$，从而满足我们前面所述的附加条件：只有价格为正且足够高，否则就不会有供给。

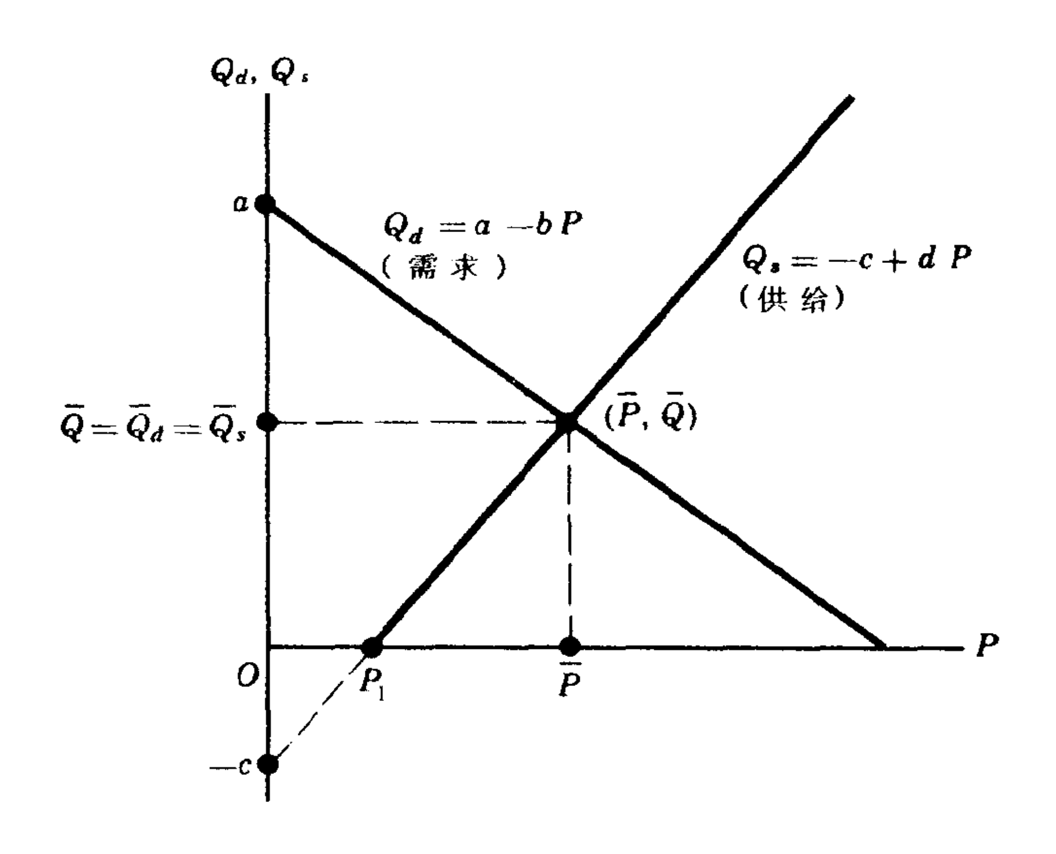



图 3.1

读者应注意到，在图 3.1 中，与通常将价格作为纵轴的画法相反，数量被作为纵轴。但这与数学上将因变量置于纵轴上的习惯是一致的。在下面的有关内容中，从厂商的角度考虑，需求曲线用于代表平均收益曲线，$AR \equiv P = f(Q_d)$，我们将把两轴的位置颠倒过来，让纵轴代表价格。 38

建模完毕，下一步就是求解，即解出三个内生变量 $Q_d$，$Q_s$ 和 $P$ 的值。解值分别以 $\bar{Q}_d$，$\bar{Q}_s$，$\bar{P}$ 表示，它们必须同时满足(3.1)中的三个方程，即将这些解值代入方程时，必须使这三个方程同时成

立。在与均衡模型相关的内容中,解值也被称作刚才所提到变量的**均衡值**。然而,因为 $\bar{Q}_d=\bar{Q}_s$,它们也可以用一个符号 $\bar{Q}$ 来表示。因此模型的均衡解可以有序偶($\bar{P}$,$\bar{Q}$)表示。当解不唯一,有几个有序偶同时满足联立方程组,则存在一个不只包含一个元素的解集。但像本例这样的线性模型中不可能出现多个均衡解的情况。

### 用变量消去法求解

解方程组的一个办法是通过代换逐步消去变量和方程。在(3.1)中,模型包括三个方程,三个变量。但考虑到由均衡条件可知 $Q_d$ 与 $Q_s$ 相等,我们可令 $Q_d=Q_s=Q$,将模型等价地改写为:

$$\begin{aligned} Q &= a - bP \\ Q &= -c + dP \end{aligned} \tag{3.2}$$

这样模型便简化为两个变量,两个方程。进而通过将(3.2)中的第一个方程代入第二个方程,模型可进一步简化为含有一个变量的一个方程。

$$a - bP = -c + dP$$

39 或者,在方程两边同时减去($a+bP$),并通乘 $-1$,有

$$(b+d)P = a + c \tag{3.3}$$

在(3.1)中直接将第二和第三个方程直接代入第一个方程,也可以得到此结果。

因 $b+d\neq 0$,所以可将(3.3)两边同时除以($b+d$),其结果便是价格的解值。

$$\bar{P} = \frac{a+c}{b+d} \tag{3.4}$$

注意，同所有解值一样，$\bar{P}$ 是完全以参数表示的，参数代表模型的给定值，所以 $\bar{P}$ 是一个确定的值。另外，$\bar{P}$ 是一个正值——它也应当是一个正值，因为模型设定四个参数均为正值。

要求出对应于 $\bar{P}$ 的均衡数量 $\bar{Q}(=\bar{Q}_d=\bar{Q}_s)$，仅需将(3.4)代入(3.2)的任一方程，并解此方程，便可以得到。例如将(3.4)代入需求函数，我们可以得到

$$\bar{Q}=a-\frac{b(a+c)}{b+d}=\frac{a(b+d)-b(a+c)}{b+d}=\frac{ad-bc}{b+d} \tag{3.5}$$

此式也是一个参数表达式。因为分母$(d+b)$为正，要使 $\bar{Q}$ 为正，则分子$(ad-bc)$也需为正。因此，要使此模型具有经济意义，还需包含额外的约束条件 $ad>bc$。

在图 3.1 中可以看出这个约束的含义。我们知道市场模型的解 $\bar{P}$和 $\bar{Q}$在图形上由供求曲线的交点决定，要使 $\bar{Q}>0$，供求曲线的交点必需位于图 3.1 横轴的上方，这就要求两曲线的斜率和纵截距的相对大小受到某些约束。根据(3.5)，给定 $b$ 和 $d$ 为正，这些约束就是 $ad>bc$。

顺便说一句，图 3.1 中供求曲线的交点与维恩图 2.2$b$ 中的交点并无概念上的不同。唯一的差别是：维恩图是位于两圆内的点；而本例则是位于两条直线上的交点。令需求曲线和供给曲线上的点集分别以 $D$ 和 $S$ 表示。则通过利用符号 $Q(=Q_s=Q_d)$，两个集合及其交点可以写成：

$$D=\{(P,Q)\mid Q=a-bP\}$$
$$S=\{(P,Q)\mid Q=-c+dP\}$$

和 $D\cap S=(\bar{P},\bar{Q})$

在本例中，交集仅包含一个元素，即有序偶$(\bar{P},\bar{Q})$。市场均衡是唯一的。

## 40 练习 3.2

---

1 给定市场模型

$$Q_d = Q_s$$
$$Q_d = 24 - 2P$$
$$Q_s = -5 + 7P$$

分别运用下述方法求出 $\bar{P}$ 和 $\bar{Q}$：

(*a*) 变量消去法；

(*b*) 运用公式(3.4)和(3.5)。(结果以分数表示，不要用小数。)

2 令需求和供给函数取下述形式：

(*a*) $Q_d = 51 - 3P$ (*b*) $Q_d = 30 - 2P$

$Q_s = 6P - 10$ $\qquad$ $Q_s = -6 + 5P$

用变量消去法求 $\bar{P}$ 和 $\bar{Q}$。(用分数表示结果，不要用小数。)

3 根据(3.5)，要使 $\bar{Q}$ 为正，一个必要条件是式$(ad - bc)$与$(b + d)$的代数符号相同。验证习题 1、2 的模型确实满足这一条件。

4 若在线性市场模型中，$(b + d) = 0$，能够利用(3.4)和(3.5)求出均衡解吗？解释为什么。

5 在线性市场模型中，若$(b + d) = 0$，对于图 3.1 中供求曲线的位置，你能得出何结论？关于均衡解，你能得出何结论？

---

## 3.3 局部市场均衡——非线性模型

在孤立的市场模型中，用二次需求函数代替线性需求函数，而

供给函数仍为线性函数。如果系数用数值而非参数，则可有下面的模型：

$$
\begin{aligned}
Q_d &= Q_s \\
Q_d &= 4 - P^2 \qquad\qquad (3.6) \\
Q_s &= 4P - 1
\end{aligned}
$$

同以往一样，这个含有三个方程的方程组可以用变量消去法（代入法）简化为一个方程：

$$4 - P^2 = 4P - 1$$

或者

$$P^2 + 4P - 5 = 0 \qquad (3.7)$$

这是一个二次方程，因为方程左侧的表达式是变量 $P$ 的二次函数。一般而言，二次方程和线性方程的主要区别在于前者有两个解值。

## 二次方程与二次函数

41

在讨论二次方程的解法以前，应对两个术语**二次方程**和**二次函数**加以明确区分。根据前面的讨论，表达式 $P^2 + 4P - 5$ 包含了一个二次**函数**，比如说 $f(P)$。因此，我们可以写成：

$$f(P) = P^2 + 4P - 5 \qquad (3.8)$$

(3.8)式规定了 $P$ 到 $f(P)$ 的一个映射规则，例如：

| $P$ | … | −6 | −5 | −4 | −3 | −2 | −1 | 0 | 1 | 2 | … |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $f(P)$ | … | 7 | 0 | −5 | −8 | −9 | −8 | −5 | 0 | 7 | … |

尽管我们在表中仅列了九个 $P$ 值，实际上函数定义域中**所有**的 $P$

值都可以列出来。也许正是由于这一原因,我们很少谈到"解"方程 $f(P)=P^2+4P-5$,因为我们一般期望"解值"仅是有限的几个数值,但这里却包括所有的 $P$ 值。当然,人们可以合理地将上表中的每一个有序偶,如(-6,7),(-5,0)等视为(3.8)的解,因为每一个这样的有序偶确实满足方程。由于有无数这样的有序偶(每一个 $P$ 值对应一个有序偶),所以,(3.8)式有无数个解。把这些有序偶联在一起绘出一条曲线,可以得到图 3.2 的抛物线。

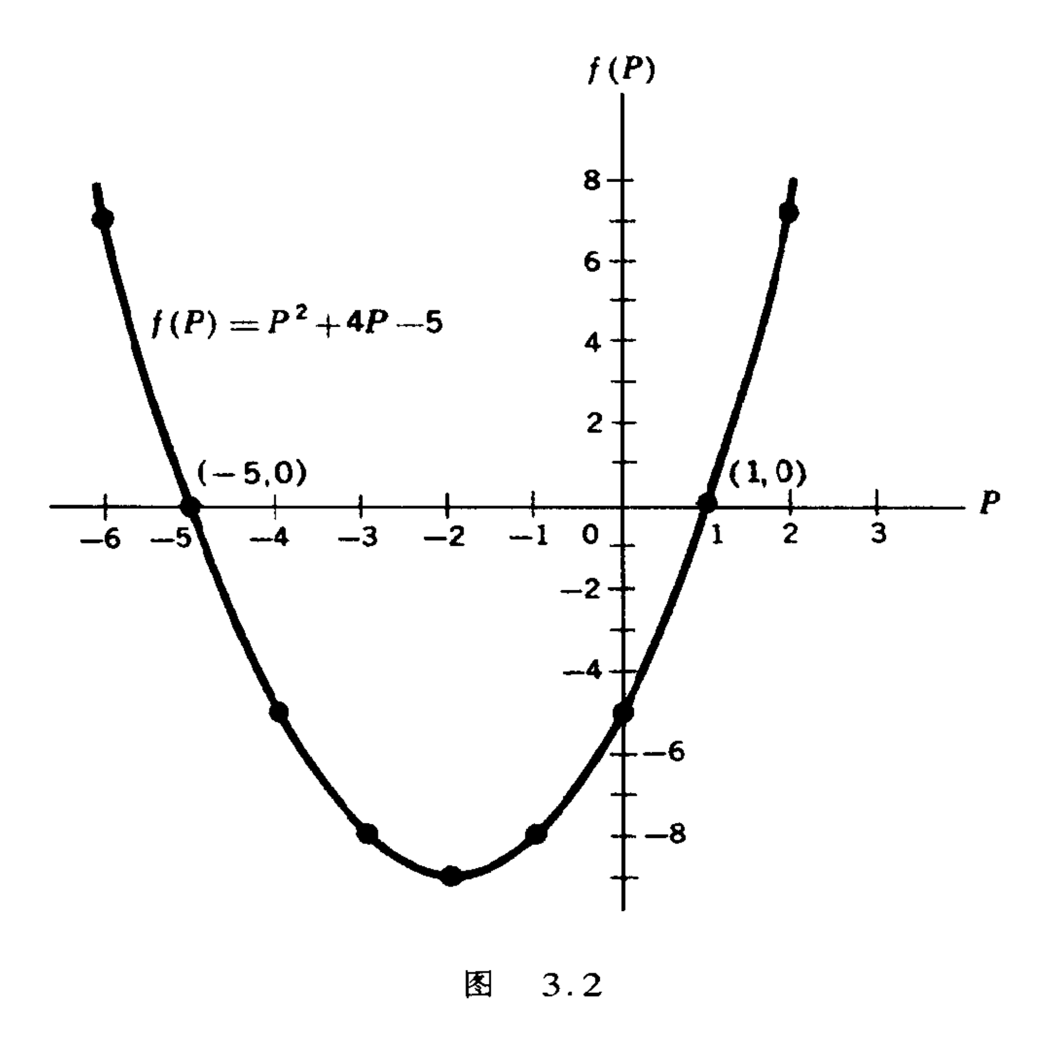



图 3.2

42 在(3.7)中,由于二次方程 $f(P)=0$,情况便产生了根本性的变化。因为变量 $f(P)$ 现在不再存在(它被赋值为零),结果二次

方程只有一个变量 $P$。[1] 既然 $f(P)$仅限取零值，那么只有有数的 $P$ 值可以满足(3.7)，并可以作为此方程的解。这些 $P$ 值也就是图 3.2 中抛物线与横轴的交点（在横轴上 $f(P)$为零）。注意，此时的解值仅仅是 $P$ 值，而非有序偶。方程的解 $P$ 值通常被称作二次方程 $f(P)=0$ 时的根，或者说当二次函数为零时的根。

在图 3.2 中，有两个这样的交点即(1,0)和(-5,0)。正如所要求的那样，有序偶的第二个元素（对应点的纵坐标）在两种情况下均表明 $f(P)$为零，而每一个有序偶的第一个元素（点的横坐标），则是解值 $P$。这里我们得到两个解：

$$\bar{P}_1 = 1 \qquad \text{和} \qquad \bar{P}_2 = -5$$

但因为负的价格被舍弃掉，仅有第一个值在经济上是可接受的。

## 二次公式

方程(3.7)已通过图解法解出，但也可以用代数法求解。一般而言，给定下述形式的二次方程

$$ax^2 + bx + c = 0 \ (a \neq 0) \tag{3.9}$$

它的两个根可通过二次公式得到

$$\bar{x}_1, \bar{x}_2 = \frac{-b \pm (b^2 - 4ac)^{1/2}}{2a} \tag{3.10}$$

其中取正号得到 $\bar{x}_1$，取负号得到 $\bar{x}_2$。

这个得到广泛运用的公式是通过“配方”的过程推导出来的。首先，(3.9)的每一项除以 $a$ 得到方程

① 这里讨论的二次函数与二次方程的区别也可以拓展到非二次的多项式的情况。因此，令三次函数为零时，就得到三次方程。

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0$$

方程两边同时减去 $c/a$，加上 $b^2/4a^2$，得到

$$x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} = \frac{b^2}{4a^2} - \frac{c}{a}$$

43 方程左侧现在是“完全平方”，因此方程可以表示成

$$(x + \frac{b}{2a})^2 = \frac{b^2 - 4ac}{4a^2}$$

或者两边开平方后

$$x + \frac{b}{2a} = \pm \frac{(b^2 - 4ac)^{1/2}}{2a}$$

最后，方程两边同时减去 $b/2a$，则可得到(3.10)的结果。

将此公式应用于(3.7)，其中 $a=1, b=4, c=-5$ 及 $x=P$，求得其根为

$$\bar{P}_1, \bar{P}_2 = \frac{-4 \pm (16+20)^{1/2}}{2} = \frac{-4 \pm 6}{2} = 1, -5$$

可将其与图 3.2 的结果加以比较验证。基于经济意义上的考虑，我们拒绝 $\bar{P}_2 = -5$。舍弃下标 1，简写为 $\bar{P} = 1$。

有了这个结果 $\bar{P} = 1$，再利用(3.6)的第二个或第三个方程，可很容易地求得均衡数量 $\bar{Q} = 3$。

## 另一种图解法

此模型图解法的一种方式已在图 3.2 中给出。但是，因为数量变量在推导二次方程时被消去，在此图中只能求出 $\bar{P}$。如果我

44 们旨在从图形中同时求出 $\bar{P}$ 和 $\bar{Q}$，必须使用以类似于图 3.1 的 $Q$ 为一轴，以 $P$ 为另一轴的图形。图 3.3 描述了这种方法，当然我

们的问题还是要求出两个点集的交集,即

$$D = \{(P, Q) \mid Q = 4 - P^2\}$$

和
$$S = \{(P, Q) \mid Q = 4P - 1\}$$

如果定义域和值域没有约束,交集应包含两个元素,即

$$D \cap S = \{(1,3), (-5, -21)\}$$

前者位于第 I 象限,后者(未画出)位于第 III 象限。但如果定义域和值域限定为非负,则只有第一个有序偶(1,3)可被接受。于是此均衡又是一个唯一的均衡。

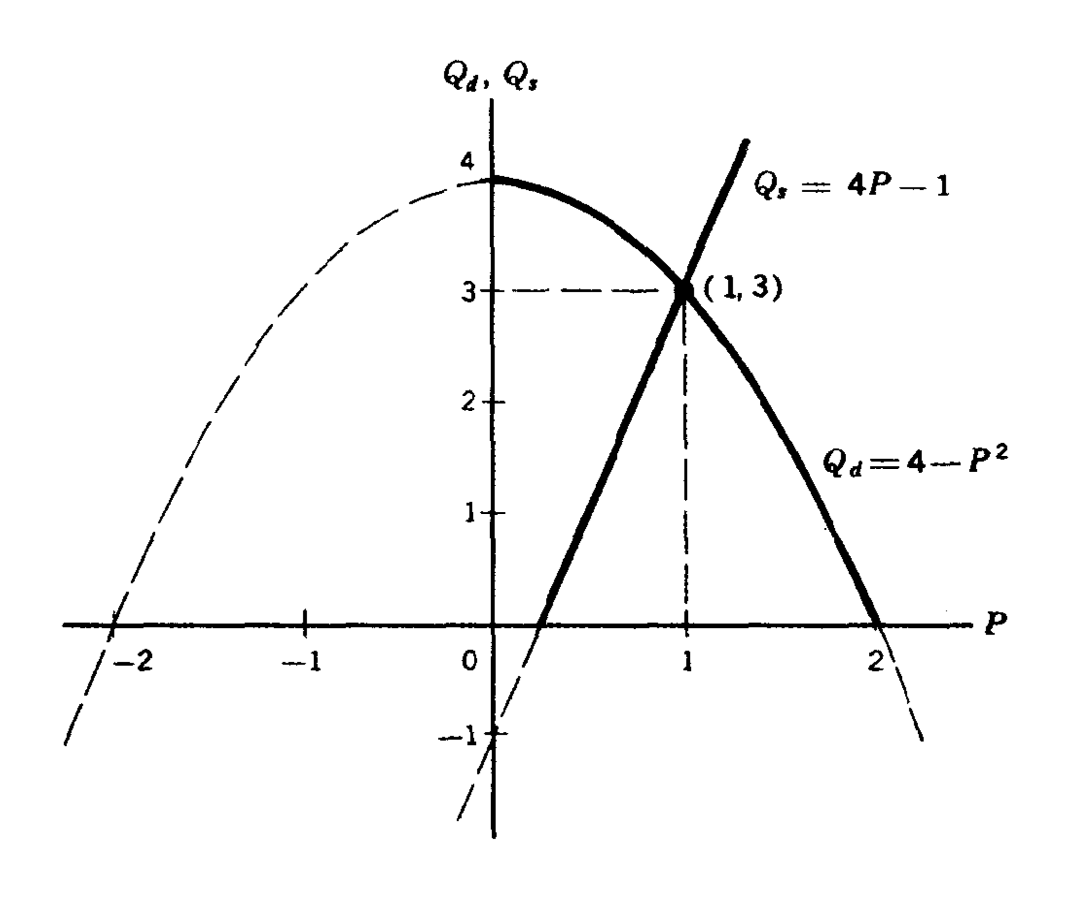



图 3.3

## 高次多项式方程

如果联立方程组不能化简成如(3.3)① 那样的线性方程，或者像(3.7)那样的二次方程，而是化简成三次(三次多项式)方程或者四次(四次多项式)方程等，求根则更为困难。一种行之有效的方法是对函数进行因式分解。例如表达式 $x^3-x^2-4x+4$ 可以写成三个因子$(x-1)$，$(x+2)$，$(x-2)$的乘积。因此，三次方程

$$x^3-x^2-4x+4=0$$

经因式分解后可写成

$$(x-1)(x+2)(x-2)=0$$

要使方程左侧的积为零，乘积中至少必有一项为零。依次令每一项为零，我们得到

$$x-1=0 \quad 或\ x+2=0 \quad 或\ x-2=0$$

这三个方程给出三次方程式的三个根，即

$$\bar{x}_1=1,\ \bar{x}_2=-2\ 及\ \bar{x}_3=2$$

当然，这种方法的关键在于找到因式分解的适当方式。但遗憾的是不存在一般的准则，因此，试错法仍是核心方法。然而，一般而言，给定 $n$ 次多项式方程 $f(x)=0$，可以预期恰有 $n$ 个根。根的求法如下：首先试找出常数 $c_1$，使得 $f(x)$可被$(x+c_1)$除。45 商 $f(x)/(x+c_1)$是一个低次即$(n-1)$次多项式函数。我们称其为 $g(x)$。由此得

$$f(x)=(x+c_1)g(x)$$

现在，试找出常数 $c_2$ 使得 $g(x)$可被$(x+c_2)$除。商 $g(x)/(x+$

---

① 方程(3.3)可视作令线性函数$(b+d)P-(a+c)$等于零的结果。

$c_2$)是一个再低一次的多项式,即($n-2$)次多项式,以 $h(x)$表示。因为 $g(x)=(x+c_2)h(x)$, 由此得:

$$f(x)=(x+c_1)g(x)=(x+c_1)(x+c_2)h(x)$$

通过重复这一过程,就可以将原来的 $n$ 次多项式 $f(x)$简化为正好是 $n$ 项的乘积:

$$f(x)=(x+c_1)(x+c_2)\cdots(x+c_n)$$

当令其为零时,将会产生 $n$ 个根。例如令第一个因式为零,得到 $\bar{x}_1=-c_1$。类似地,由其它各个因式可得 $\bar{x}_2=-c_2, \bar{x}_3=-c_3$ 等等。使用下标 $i$,可以将这些结果简洁地表示为

$$\bar{X}_i=-C_i \qquad (i=1,2,\cdots,n)$$

尽管这里只写出一个方程,但下标 $i$ 可以取 $n$ 个不同的值,这意味着包含了所有 $n$ 个方程。所以下标提供了一种非常简洁的表达方式。

## 练习 3.3

---

1 用图解法求出下列函数为零时 $x$ 的值。

(a) $f(x)=x^2-7x+10$

(b) $g(x)=2x^2-4x-16$

2 用二次公式法求解上述问题。

3 运用因式分解法求解下列多项式:

(a) $P^2+4P-5=0$ [参见(3.7)]

(b) $x^3+2x^2-4x-8=0$

(c) $x^3-7x^2+14x-8=0$

(d) $x^3-3x^2-4x=0$

4　求出一个根为 7，-2 和 5 的三次方程。

5　求出下列每一模型的均衡解：

(a) $Q_d = Q_s$　　(b) $Q_d = Q_s$

$Q_d = 3 - P^2$　　$Q_d = 8 - P^2$

$Q_s = 6P - 4$　　$Q_s = P^2 - 2$

6　市场均衡条件 $Q_d = Q_s$ 也常以另一种形式 $Q_d - Q_s = 0$ 来表示，其经济解释为“超额需求为零”。(3.7)式代表均衡条件的后一种形式吗？如果不是，给出(3.7)式的恰当的经济解释。

---

## 3.4 一般市场均衡

46　上两节讨论了孤立的市场模型，在那里商品的 $Q_d$ 和 $Q_s$ 仅仅是该商品价格的函数。然而在现实世界中，没有一种商品是这样孤立存在的，每一种商品都有许多替代品和互补品。因此，对一个商品需求函数的更为实际的描述不但应考虑到商品自身价格的影响，还应考虑到大部分(如果不能做到全部)相关产品价格的影响。供给函数也应如此。但是，一旦其它商品价格被纳入考虑范围，模型的结构必须加以扩大，以便能求出其它商品价格的均衡值。因此，多种商品的价格和数量变量必须一并以内生变量纳入模型。

在孤立市场模型中，均衡条件仅包含一个方程 $Q_d = Q_s$，或 $E = Q_d - Q_s = 0$，其中 $E$ 代表超额需求。当同时考虑几种相互关联的商品时，均衡要求模型中的每一个商品都不存在超额需求。因为只要有一种商品存在超额需求，该商品的价格调整就会影响到其它商品的需求数量和供给数量，从而导致所有商品的价格变化。总之，$n$ 种商品市场模型的均衡条件将包含 $n$ 个方程，每个方

程代表一种商品,其形式为:

$$E_i = Q_{di} - Q_{si} = 0 \ (i = 1, 2, \cdots, n) \qquad (3.11)$$

如果存在均衡解,它将是一个价格 $\bar{P}_i$ 集合和数量 $\bar{Q}_i$ 集合,使得均衡条件中的所有 $n$ 个方程同时得到满足。

## 两种商品市场模型

为了说明问题,我们讨论一种仅包含两种相互关联的商品的简单模型。为简化起见,两种商品的需求函数和供给函数均假设为线性的。用参数形式,这种模型可以写成

$$\begin{aligned} &Q_{d1} - Q_{s1} = 0 \\ &Q_{d1} = a_0 + a_1 P_1 + a_2 P_2 \\ &Q_{s1} = b_0 + b_1 P_1 + b_2 P_2 \\ &Q_{d2} - Q_{s2} = 0 \\ &Q_{d2} = \alpha_0 + \alpha_1 P_1 + \alpha_2 P_2 \\ &Q_{s2} = \beta_0 + \beta_1 P_1 + \beta_2 P_2 \end{aligned} \qquad (3.12)$$

其中系数 $a$ 和 $b$ 属于第一种商品的需求和供给函数,$\alpha$ 和 $\beta$ 属于 47 第二种商品的需求和供给函数。我们没有设定系数的符号,但在分析过程中为使结果具有经济意义,要施加某些限制以作为先决条件。在后面的数值系数的例子中,对于某些系数的特定符号还要予以说明。

作为求解此模型的第一步,我们再次利用变量消去法。通过将第二个、第三个方程代入第一个方程(描述第一个商品的方程),第五、第六个方程代入第四个方程(描述第二个商品的方程),模型可以简化为含有两个变量的两个方程:

$$
\begin{aligned}
&(a_0 - b_0) + (a_1 - b_1)P_1 + (a_2 - b_2)P_2 = 0 \\
&(\alpha_0 - \beta_0) + (\alpha_1 - \beta_1) P_1 + (\alpha_2 - \beta_2) P_2 = 0
\end{aligned}
\tag{3.13}
$$

将需求和供给函数代入两个均衡条件方程之后，就得到了(3.11)式的两种商品的形式。

尽管这是一个仅含有两个方程的简单方程组，但所包含的参数达 12 个，如果不引入一些简写符号，代数处理将是极其麻烦的。因此，我们定义简写符号：

$$
\begin{aligned}
c_i &\equiv a_i - b_i \\
\gamma_i &\equiv \alpha_i - \beta_i
\end{aligned}
\qquad (i = 0,1,2)
$$

将 $c_0$ 和 $\gamma_0$ 移至方程等号右侧后，(3.13)式变成：

$$
\begin{aligned}
c_1 P_1 + c_2 P_2 &= - c_0 \\
\gamma_1 P_1 + \gamma_2 P_2 &= - \gamma_0
\end{aligned}
\tag{3.13'}
$$

通过进一步使用变量消去法，此式可解。由第一个方程可得 $P_2 = -(c_0 + c_1 P_1)/c_2$。将其代入第二个方程并解之，我们得到

$$
\bar{P}_1 = \frac{c_2\gamma_0 - c_0\gamma_2}{c_1\gamma_2 - c_2\gamma_1} \tag{3.14}
$$

注意 $\bar{P}_1$ 是完全以模型中的数据(参数)表示的，作为解值，它也应当如此。通过类似的过程，求得第二种商品的均衡价格为：

$$
\bar{P}_2 = \frac{c_0\gamma_1 - c_1\gamma_0}{c_1\gamma_2 - c_2\gamma_1} \tag{3.15}
$$

为使这两个解值有意义，对模型需施加某些限制。首先，因为用零去除没有意义，我们要求(3.14)和(3.15)中的相同的分母不为零，即 $c_1\gamma_2 \neq c_2\gamma_1$。第二，为保证 $\bar{P}$ 为正，分子、分母的符号应相同。



求出均衡价格后，通过将(3.14)和(3.15)代入(3.12)的第二(或第三)个方程和第五(或第六)个方程，可以求出均衡数量 $\bar{Q}_1$

和 $\bar{Q}_2$，其解值自然也应以参数表示。（实际计算过程留给读者作为练习。）

**数值例子**

假设需求和供给函数取下述数值形式：

$$\begin{aligned} Q_{d1} &= 10 - 2P_1 + P_2 \\ Q_{s1} &= -2 + 3P_1 \\ Q_{d2} &= 15 + P_1 - P_2 \\ Q_{s2} &= -1 + 2P_2 \end{aligned} \tag{3.16}$$

何谓其均衡解？

在回答此问题前，我们先考察一下数值系数。对于每一商品，$Q_{si}$ 仅取决于 $P_i$，但 $Q_{di}$ 是两种商品价格的函数。注意，尽管在 $Q_{d1}$ 中，$P_1$ 的系数为负，但正如我们期望的那样，$P_2$ 的系数为正。$P_2$ 上升使 $Q_{d1}$ 增加这一事实意味着这两种商品互为替代品。在 $Q_{d2}$ 中 $P_1$ 的作用也有类似的解释。

运用上述系数，简写符号 $C_i$ 和 $\gamma_i$ 将取以下值：

$$c_0 = 10 - (-2) = 12 \quad c_1 = -2 - 3 = -5 \quad c_2 = 1 - 0 = 1$$

$$\gamma_0 = 15 - (-1) = 16 \quad \gamma_1 = 1 - 0 = 1 \quad \gamma_2 = -1 - 2 = -3$$

直接将其代入(3.14)和(3.15)，我们得到

$$\bar{P}_1 = \frac{52}{14} = 3\frac{5}{7} \text{ 和 } \bar{P}_2 = \frac{92}{14} = 6\frac{4}{7}$$

再进一步将 $\bar{P}_1$ 和 $\bar{P}_2$ 代入(3.16)，有

$$\bar{Q}_1 = \frac{64}{7} = 9\frac{1}{7} \text{ 和 } \bar{Q}_2 = \frac{85}{7} = 12\frac{1}{7}$$

于是所有均衡值均如所要求的那样，都是正值。为了保留 $\bar{P}_1$ 和

$\bar{P}_2$ 的精确值以便在进一步计算 $\bar{Q}_1$ 和 $\bar{Q}_2$ 时使用，以分数而非小数表示 $\bar{P}_1$ 和 $\bar{P}_2$ 是明智的。

我们能用图解法解出均衡价格吗？答案是肯定的。由(3.13)，很显然两商品模型可以由具有两个变量 $\bar{P}_1$、$\bar{P}_2$ 的两个方程概括而成。运用已知数值系数，在 $P_1P_2$ 坐标平面中可绘出两个方程的曲线，两条曲线的交点则可准确地确定 $\bar{P}_1$ 和 $\bar{P}_2$。

## 49 *n* 种商品的情况

上面对多种商品市场模型的讨论仅限于两种商品的情况，但很明显我们已由局部均衡分析向一般均衡分析转变。当更多的商品进入模型时，变量和方程的数量也随之增加，方程也会变得更长和复杂。如果综合市场模型包括一个经济的所有商品，则此模型便是瓦尔拉斯一般均衡模型。在此模型中，每一商品的超额需求被视为该经济中所有商品价格的函数。

当然，当某些商品的价格对某个特殊商品的超额需求不产生作用时，其系数可取值为零。例如在钢琴的超额需求函数中，花生价格的系数可以取零值。但一般而言，具有 $n$ 种商品的需求函数和供给函数可以表述成如下形式(使用 $Q_{di}$ 和 $Q_{si}$ 作为函数符号以代替 $f$ 和 $g$)：

$$\begin{aligned} Q_{di} &= Q_{di}(P_1, P_2, \cdots, P_n) \\ Q_{si} &= Q_{si}(P_1, P_2, \cdots, P_n) \end{aligned} \qquad (i = 1, 2, \cdots, n) \qquad (3.17)$$

根据下标可知，这两个方程代表了模型所包含的所有 $2n$ 个方程。(这些方程不必都是线性的。)而且，均衡条件本身由 $n$ 个方程的集合组成：

$$Q_{di} - Q_{si} = 0 \; (i = 1, 2, \cdots, n) \qquad (3.18)$$

把(3.17)和(3.18)加在一起,便成为完整的模型,共含有 $3n$ 个方程。

然而,把(3.17)代入(3.18),模型可以简化为仅含有 $n$ 个方程的联立方程组:

$$Q_{di}(P_1,P_2,\cdots,P_n) - Q_{si}(P_1,P_2,\cdots,P_n) = 0 \quad (i = 1,2,\cdots,n)$$

此外,由于 $E_i = Q_{di} - Q_{si}$,其中 $E_i$ 必然也是 $n$ 种商品价格的函数,上述方程组还可以写成

$$E_i(P_1,P_2,\cdots,P_n) = 0 \quad (i = 1,2,\cdots,n)$$

如果确实有解,联立求解,$n$ 个方程将决定 $n$ 个价格 $\bar{P}_i$,而 $\bar{Q}_i$ 则可从需求函数或供给函数推导出来。

## 一般方程组的解

若一模型具有数值系数,如(3.16)式,则变量均衡值也将是数值的。在较高的一般性水平上,如果一模型像(3.12)那样,以参常数表示,则均衡值也由参数表示,并如(3.14)、(3.15)那样,以“公式”形式出现。但如果为了获得更广泛的一般性,比如像(3.17)那样,对模型中函数的形式也不作设定,那么,解值的表示方式也必然有更为广泛的一般性。

50

根据参数模型中的经验,我们知道解值总是以参数表示的。因此,对于一般函数模型,比如含有 $m$ 个参数$(a_1,a_2,\cdots,a_m)$的一般函数模型($m$ 不一定等于 $n$),可以预期 $n$ 种均衡价格取如下一般分析形式:

$$\bar{P}_i = \bar{P}_i(a_1,a_2,\cdots,a_m) \quad (i = 1,2,\cdots n) \qquad (3.19)$$

此式是对每一变量(这里是价格)的解值是模型所有参数集合的函

数这一结论的符号表述。因为这是一个极一般化的描述，它确实没有给出解的详细信息。但是正如我们在后面的章节中将要看到的那样，在对此类问题的一般分析与处理方式中，即使这种对解的不含信息的表达方式也是有益的。

写出这样的解是一个轻松的工作，但还存在一个意外的困难：当且仅当只存在唯一解时，表达式(3.19)才有意义；对于每一个价格 $P_i$，当且仅当(3.19)式有意义时，我们才能把 $m$ 元数组($a_1$, $a_2$, …, $a_m$)映射到每一个确定的值。但遗憾的是，我们并无先验的理由认为每一个模型会自动产生一个唯一解。在这方面，需要强调的是，“计算方程和未知数的个数”不足以作为一种检验方法。下面一些简单的例子足以证明，方程数量和未知数(内生变量)相等不足以保证唯一解的存在。

考察下面三个联立方程组：

$$\begin{aligned} x + y &= 8 \\ x + y &= 9 \end{aligned} \tag{3.20}$$

$$\begin{aligned} 2x + \ \ y &= 12 \\ 4x + 2y &= 24 \end{aligned} \tag{3.21}$$

$$\begin{aligned} 2x + 3y &= 58 \\ y &= 18 \\ x + y &= 20 \end{aligned} \tag{3.22}$$

在(3.20)中，尽管两个未知数确实通过两个方程联系起来，但方程组无解。因为这两个方程是不相容的，因为 $x$ 加 $y$ 若等于 8，则二者之和不可能同时还等于 9。(3.21)则给出了两变量两方程的另

一种情况:这两个方程是函数相关的,即可以从一个函数中推导出另一个函数,或一个函数中隐含着另一个函数。(这里,第二个方程为第一个方程的二倍。)因而,一个方程是多余的,可从方程组中去掉,结果只留下含有两个未知数的一个方程,其解可以是 $y=12-2x$ [51],由它不能得到一个唯一的有序偶($\bar{x}$,$\bar{y}$),而是一组无穷的数,包括(0,12)、(1,10)、(2,8)等等,所有这些均使方程成立。最后一个例子(3.22)则给出了方程数多于未知数个数的情况,但有序偶(2,18)确实构成了此方程组的唯一解。其原因是,由于方程组中存在函数相关(第一个方程等于第二个方程与第三个方程二倍之和),我们实际上只有两个变量,两个独立的相容的方程。

这些简单的例子足以表明**相容性**和**函数非相关性**作为应用计算方程和未知数个数方法的先决条件的重要性。一般而言,应用这种方法应确保:(1)变量满足模型中任一方程必须满足模型中其它方程;(2)模型中无多余方程。如在(3.17)中,无疑可以假设 $n$ 个需求函数和 $n$ 个供给函数是彼此无关的,因为每个函数均有不同的来源:每一需求函数源于一组消费者的决策,而每一供给函数源于一组厂商的决策。因此,每一函数用于描述市场状况的一个侧面,没有一个函数是多余的。我们还可以假设(3.17)具有相容性。此外,均衡条件方程(3.18)也是非相关的,且可假设它们是相容的。因此,写成如(3.19)那样的分析解一般认为是有道理的。①

① 这实质上是瓦尔拉斯分析一般市场均衡存在问题所采用的方法。在现代文献中,可以看到在某些假定条件下,对竞争市场均衡存在的复杂的数学证明。但他们所运用的数学是深奥的。其中最易理解的可能是罗伯特·多夫曼,保罗·萨缪尔森和罗伯特·索洛在《线性规划与经济分析》中的证明(麦克劳—希尔出版公司,纽约,1958 年版,第 13 章),读者在学完本书第六编后可以阅读该书。

对于联立方程组,有系统方法检验其唯一(或定值)解的存在。对于线性模型,还可以运用将于第五章中介绍的行列式的概念。在非线性模型情况下,这种检验要运用"偏导数"的知识以及雅可比行列式这类特殊类型的行列式。后者我们将要在第7、8章中讨论。

### 练习 3.4

---

1 逐步求解(3.13′),由此验证(3.14),(3.15)的结果。

2 用模型(3.12)中的原始参数重写(3.14)和(3.15)式。

3 两商品市场模型的需求和供给函数如下:

$$Q_{d1} = 18 - 3P_1 + P_2 \qquad Q_{d2} = 12 + P_1 - 2P_2$$

$$Q_{s1} = -2 + 4P_1 \qquad Q_{s2} = -2 + 3P_2$$

求 $\bar{P}_i, \bar{Q}_i (i = 1, 2)$。(使用分数,不要使用小数。)

---

52

## 3.5 国民收入分析中的均衡

尽管静态分析迄今为止仅限于讨论各类市场模型,比如线性的和非线性的,一种商品和多种商品,特殊的和一般的等等,它当然还可以应用于经济分析的其它领域。我们引用熟悉的凯恩斯国民收入模型作为一个简单的例子:

$$\begin{aligned} Y &= C + I_0 + G_0 \\ C &= a + bY \end{aligned} \qquad (a > 0, 0 < b < 1) \qquad (3.23)$$

其中 $Y$ 和 $C$ 分别表示内生变量国民收入和消费支出,$I_0$ 和 $G_0$ 分

别表示外生决定的投资和政府支出。第一个方程是均衡条件(国民收入等于总支出)。第二个方程即消费函数,是行为方程。消费函数中的两个参数 $a$ 和 $b$ 分别代表自发消费支出和边际消费倾向。

很明显,这两个方程中包含两个内生变量,它们既不存在函数相关,也非互不相容。因此,我们可以求出以参数 $a$、$b$ 和外生变量 $I_0$、$G_0$ 表示的收入与消费支出的均衡值 $\bar{Y}$ 和 $\bar{C}$。

将(3.23)的第二个方程代入第一个方程,可将其简化为仅包含一个变量 $Y$ 的一个方程:

$$Y = a + bY + I_0 + G_0$$

或

$$(1 - b)Y = a + I_0 + G_0$$

因而 $Y$ 的解值(均衡国民收入)为

$$\bar{Y} = \frac{a + I_0 + G_0}{1 - b} \tag{3.24}$$

需要指出的是,此式完全是以模型给定的数据,即参数和外生变量表示的。将(3.24)代入(3.23)的第二个方程,则得到均衡的消费支出水平:

$$\begin{aligned}\bar{C} = a + b\bar{Y} &= a + \frac{b(a + I_0 + G_0)}{1 - b} \\ &= \frac{a(1 - b) + b(a + I_0 + G_0)}{1 - b} = \frac{a + b(I_0 + G_0)}{1 - b}\end{aligned} \tag{3.25}$$

此式也是完全以给定参数表示的。

$\bar{Y}$ 和 $\bar{C}$ 的分母均具有表达式$(1 - b)$,因此,限定 $b \neq 1$ 是必要的,以避免被零除。因为边际消费倾向 $b$ 已被假设为一个正分

53数，所以 $b \neq 1$ 的限制自然满足。因为要使 $\bar{Y}$ 和 $\bar{C}$ 为正，则(3.24)、(3.25)的分子也必须为正。由于外生支出 $I_0$ 和 $G_0$ 通常为正，参数 $a$（消费函数的纵截距）为正，所以分子表达式的符号也就确定了。

作为对我们计算的一项检验，我们可以把 $\bar{C}$ 表达式(3.25)加上$(I_0 + G_0)$，看此和是否与 $\bar{Y}$ 表达式(3.24)相等。如果相等，则 $\bar{C}$ 和 $\bar{Y}$ 值满足均衡条件，此解有效。

很明显，此模型是极为简单、粗糙的。但其它复杂性和完备性不同的国民收入决定模型，也可以建立起来。而建立和分析这些模型的原理同前面的讨论是一致的，所以我们不在此作进一步说明。货币市场和产品市场同时处于均衡状态的更为综合的国民收入模型，将在第 8 章第 6 节中讨论。

## 练习 3.5

---

1 给定下列模型

$$Y = C + I_0 + G_0$$
$$C = a + b(Y - T) \qquad (a > 0,\ 0 < b < 1)\ [T:\text{税收}]$$
$$T = d + tY \qquad (d > 0,\ 0 < t < 1)\ [t:\text{所得税率}]$$

($a$) 有几个内生变量？

($b$) 求解 $\bar{Y}$，$\bar{T}$，$\bar{C}$。

2 令国民收入模型为

$$Y = C + I_0 + G$$
$$C = a + b(Y - T_0) \qquad (a>0,\ 0<b<1)$$
$$G = gY \qquad (0<g<1)$$

($a$) 识别内生变量？

($b$) 解释参数 $g$ 的经济意义;

($c$) 求均衡国民收入;

($d$) 要使解存在,对参数要施加何种限制。

3 由下述模型求解 $\bar{Y}$ 和 $\bar{C}$:

$$Y = C + I_0 + G_0$$

$$C = 25 + 6Y^{1/2}$$

$$I_0 = 16$$

$$G_0 = 14$$

54

# 第4章　线性模型与矩阵代数

仅有一种商品的模型(3.1)的解分别以(3.4)和(3.5)表示。尽管解中包含若干参数,但还是相对简单的。随着越来越多的商品纳入模型,解的公式变得越来越庞杂,难于应用。这就是我们之所以寻求简化表达的原因。即使对于两种商品的情况,我们仍寻求简化的表达方式,以使其解(3.14)和(3.15)能以相对简洁的方式表示出来。我们并不谋求解决三种或四种商品的模型,即使是线性模型,这主要是因为我们还没有掌握处理复杂的联立方程组的合适方法。这种方法可在矩阵代数中找到,这是本章和下一章的主题。

矩阵代数可使我们得以处理很多问题。首先,它提供了一种描写方程组——即使是非常庞大的方程——的简洁方式。其次,它引出一种通过估计行列式(与矩阵密切相关的一个概念)来检验解的存在的一种方法。第三,它提供一种求解的方法(若解存在的话)。因为方程组不仅在静态分析中要遇到,而且在比较静态分析、动态分析和最优化问题中也要遇到,所以读者将会看到矩阵代数在本书以后各章中几乎都有应用。

然而在本章开始还要提到一个小小的“困难”:矩阵代数仅可应用于线性方程组。当然,线性方程能在多大程度上描述实际经济关系,取决于所研究问题的关系的性质。在很多情况下,即使由于线性假设可能使真实性蒙受某些损失,但假设的线性关系仍能

与实际的非线性关系产生足够充分的近似，从而保证其使用。在另外一些情况下，对非线性关系采用分段线性近似的方法可以提高近似的精确度，如图 4.1 所示。在图 4.1 中，若实曲线表示实际 55 非线性关系，一个单一线性近似取实直线形式，由此可见直线上的某些点与曲线有极大的偏离。但若我们将定义域分为三个区域 $r_1, r_2, r_3$，在每一区域内我们可以得到更精确的线性近似（折直线）。

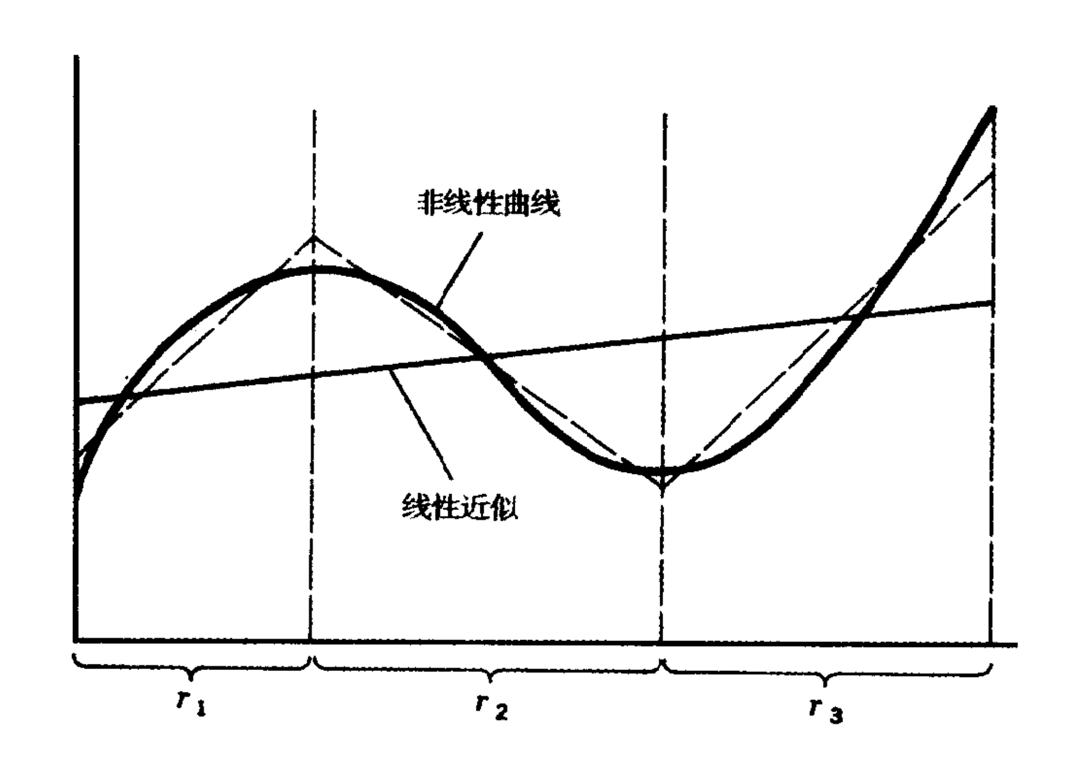



图 4.1

还有一种情况，尽管模型中保留非线性关系，我们可以通过变量变换，以得到可以处理的线性关系。例如，非线性函数

$$y = ax^b$$

可通过对两边取对数，变换成函数

$$\log y = \log a + b\log x$$

它是一个具有两个变量的线性函数（$\log y$ 和 $\log x$）。（对数将在

第 10 章中详细讨论。)

总之,在某些情况下经济学中经常采用的线性假设可能是非常合理和有意义的。通过这样的解释,我们便可以着手学习矩阵代数。

## 4.1 矩阵与向量

56 两种商品市场模型(3.12)在消去数量变量之后,可以写成如(3.13′)那样的由两个线性方程组成的方程组,

$$c_1P_1 + c_2P_2 = -c_0$$
$$\gamma_1P_1 + \gamma_2P_2 = -\gamma_0$$

此处参数 $c_0$ 与 $\gamma_0$ 出现在等号右边。一般而言,含有 $n$ 个变量 $(x_1, x_2, \cdots, x_n)$ 的 $m$ 个线性方程组成的方程组可以整理成这种形式:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= d_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= d_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= d_m \end{aligned} \tag{4.1}$$

在(4.1)中,$x_1$ 仅出现于最左侧的一列。一般而言,$x_j$ 仅出现于等号左侧第 $j$ 列。双下标参数符号 $a_{ij}$ 表示第 $i$ 个方程第 $j$ 个变量的系数。例如 $a_{21}$ 表示第 2 个方程第 1 个变量 $x_1$ 的系数。而参数 $d_i$,它不与任何一个变量相联,表示第 $i$ 个方程的常数项。例如,$d_1$ 表示第 1 个方程的常数项。因此,所有的下标都表示变量和参数在(4.1)中的特定位置。

## 作为阵列的矩阵

在方程组(4.1)中主要有三种类型的成份。第一种为系数 $a_{ij}$ 的集合;第二种是变量 $x_1,\cdots,x_n$ 的集合;最后一种则是常数项 $d_1,\cdots,d_m$ 的集合。若我们将此三个集合排成三个矩阵列,并分别以 $A$,$x$ 和 $d$(无下标)标示,则我们有

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad x=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad d=\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{bmatrix} \tag{4.2}$$

作为一个简单的例子,给定线性方程组

$$\begin{aligned} 6x_1+3x_2+x_3&=22 \\ x_1+4x_2-2x_3&=12 \\ 4x_1-x_2+5x_3&=10 \end{aligned} \tag{4.3}$$

我们可以写出

$$A=\begin{bmatrix} 6 & 3 & 1 \\ 1 & 4 & -2 \\ 4 & -1 & 5 \end{bmatrix} \quad x=\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad d=\begin{bmatrix} 22 \\ 12 \\ 10 \end{bmatrix} \tag{4.4}$$

(4.2)或(4.4)中三个阵列中的每一个阵列,均构成一个**矩阵**。

矩阵定义为数字、参数或变量的矩形阵列。阵列中的成员,也指矩阵中的**元素**,通常如(4.2)式,以括弧括之,有时也用小括号或双竖线围之。注意,在矩阵 $A$(方程组系数矩阵)中,元素不以逗号分开,而仅以空格分开。作为一种简写方式,矩阵 $A$ 中的阵列还可以更简单地写成

$$A = [a_{ij}] \quad \begin{pmatrix} i = 1,2,\cdots,m \\ j = 1,2,\cdots,n \end{pmatrix}$$

由于矩阵中每一元素的位置由下标明确确定，所以每一矩阵均是一个有序集合。

### 作为特殊矩阵的向量

矩阵中的行数和列数一起定义为矩阵的维数。因为(4.2)中的矩阵 $A$ 包括 $m$ 行，$n$ 列，所以称其为 $m \times n$ 维矩阵(读作“$m$ 乘 $n$”)。记住行数总是在列数的前面是重要的。这与 $a_{ij}$ 下标的顺序是一致的。当 $m = n$ 时的特殊情况，矩阵被称作方阵；所以(4.4)中的矩阵 $A$ 为 $3 \times 3$ 方阵。

有些矩阵可能仅包含一列，如(4.2)或者(4.4)中的 $x$ 和 $d$。这样的矩阵有一个特别的名称列向量。在(4.1)中，$x$ 的维数为 $n \times 1$，$d$ 的维数为 $m \times 1$。在(4.4)中，$x$ 和 $d$ 的维数均为 $3 \times 1$。如果我们将变量 $x_j$ 排成横的阵列，则会得到 $1 \times n$ 阶矩阵，我们称其为行向量。我们通常使用撇号：

$$x' = [x_1 \ x_2 \ \cdots \ x_n]$$

以使行向量和列向量区别开来。读者应注意到无论行向量还是列向量，只是一个有序 $n$ 元数组，因此它可以解释为 $n$ 维空间中的一个点。进而，$m \times n$ 维矩阵 $A$ 可以解释为 $m$ 个行向量的有序集合，或者 $n$ 个列向量的有序集合。这种思想以后还要继续应用。

一个更令人感兴趣的问题是：矩阵这个概念如何能像我们承诺的那样，以更简洁的方式表示方程组。运用(4.4)定义的矩阵，我们可以将方程组(4.3)表示成

$$Ax = d$$

事实上，若 $A, x, d$ 的意义与(4.2)相同，则一般方程组(4.1)也可以写成 $Ax = d$。所以矩阵概念的简洁性是毫无疑问的。

然而，方程 $Ax = d$ 至少会引起两个问题：我们如何将两个矩阵 $A$ 和 $x$ 相乘？因为矩阵包括整个的数块，我们熟悉的对单个数定义的代数运算难以直接应用，所以需要新的运算规则。

## 练习 4.1

58

---

1 按(4.1)方式改写方程组(3.1)，并证明：若三个变量的顺序重排为 $Q_d, Q_s$ 和 $P$，则系数矩阵将为：

$$\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & b \\ 0 & 1 & -d \end{bmatrix}$$

如何写出常数向量？

2 变量按下列顺序 $Q_{d1}, Q_{s1}, Q_{d2}, Q_{s2}, P_1, P_2$，按(4.1)方式改写方程组(3.12)。写出系数矩阵，变量向量和常数向量。

---

# 4.2 矩阵运算

作为基础，我们先定义矩阵"相等"。两个矩阵 $A = \{a_{aj}\}$ 和 $B = \{b_{ij}\}$，当且仅当二者有相同的维数，且阵列中对应位置的元素相同，我们才称二者相等，换句话说，对于所有的 $i$ 和 $j$，当且仅当 $a_{ij} = b_{ij}$，则 $A = B$。因此，我们有此例：

$$\begin{bmatrix} 4 & 3 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 3 \\ 2 & 0 \end{bmatrix} \neq \begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix}$$

又如，若$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \end{bmatrix}$，这意味着 $x=7$，且 $y=4$。

## 矩阵的加法和减法

当且仅当两个矩阵具有相同维数时，二者才可以相加。当这种维数要求得到满足时，则称矩阵是**可相加的**。在此情况下，矩阵 $A=[a_{aj}]$和 $B=[b_{ij}]$相加定义为每一对对应元素相加。

**例 1**

$$\begin{bmatrix} 4 & 9 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} 4+2 & 9+0 \\ 2+0 & 1+7 \end{bmatrix} = \begin{bmatrix} 6 & 9 \\ 2 & 8 \end{bmatrix}$$

59

**例 2**

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \\ a_{31}+b_{31} & a_{32}+b_{32} \end{bmatrix}$$

一般情况下，我们可将这一规则表述成

$$[a_{ij}] + [b_{ij}] = [c_{ij}] \qquad 其中\ c_{ij} = a_{ij} + b_{ij}$$

注意，和矩阵$[c_{ij}]$与其分量矩阵$[a_{ij}]$和$[b_{ij}]$的维数一定相同。

类似地，两个矩阵的减法运算 $A-B$ 的条件可以定义为，当且仅当 $A$ 与 $B$ 具有相同维数时，才可以进行减法运算。运算产生结果

$$[a_{ij}] - [b_{ij}] = [d_{ij}] \qquad 其中\ d_{ij} = a_{ij} - b_{ij}$$

**例 3**

$$\begin{bmatrix} 19 & 3 \\ 2 & 0 \end{bmatrix} - \begin{bmatrix} 6 & 8 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 19-6 & 3-8 \\ 2-1 & 0-3 \end{bmatrix} = \begin{bmatrix} 13 & -5 \\ 1 & -3 \end{bmatrix}$$

减法运算 $A-B$ 还可以视为矩阵 $A$ 与另一矩阵$(-1)B$ 相加。但这又引出了矩阵如何与一个单个数（这里为 $-1$）相乘的问题。

## 标量乘法

一个数,或者用矩阵代数的语言来说,一个标量与一个矩阵相乘,就是用这个给定标量与该矩阵的每一个元素相乘。

**例 4**

$$7\begin{bmatrix}3 & -1\\ 0 & 5\end{bmatrix}=\begin{bmatrix}21 & -7\\ 0 & 35\end{bmatrix}$$

**例 5**

$$\frac{1}{2}\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}=\begin{bmatrix}\frac{1}{2}a_{11} & \frac{1}{2}a_{12}\\ \frac{1}{2}a_{21} & \frac{1}{2}a_{22}\end{bmatrix}$$

由这些例子,我们可以明了标量这个术语的合理性了。因为它按一定的乘数"按比例扩大(或缩小)"矩阵。当然,标量也可以是负数。

**例 6**

$$-1\begin{bmatrix}a_{11} & a_{12} & d_1\\ a_{21} & a_{22} & d_2\end{bmatrix}=\begin{bmatrix}-a_{11} & -a_{12} & -d_1\\ -a_{21} & -a_{22} & -d_2\end{bmatrix}$$

注意,若左侧的矩阵表示下列联立方程组

60

$$a_{11}x_1+a_{12}x_2=d_1$$

$$a_{21}x_1+a_{22}x_2=d_2$$

中的系数和常数项,则以标量 -1 相乘等价于在方程两侧同时以 -1 相乘,因而改变了方程组每一项的符号。

## 矩阵乘法

尽管标量可与任意维数的矩阵相乘,但两个矩阵能否相乘则

要视不同的维数要求能否得到满足而定。

给定两个矩阵 $A$ 和 $B$,假设我们要求乘积 $AB$。矩阵可相乘的条件是 $A$ 矩阵(表达式 $AB$ 中的"前"一个矩阵)的列维数,一定等于 $B$(后一矩阵)的行维数。例如,若

$$\underset{(1\times2)}{A} = \begin{bmatrix} a_{11} & a_{12} \end{bmatrix} \quad \underset{(2\times3)}{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} \tag{4.5}$$

乘积 $AB$ 符合定义,因 $A$ 具有两列,$B$ 有两行,二者恰好相等。[①] 将表示矩阵 $A$ 的维数(1×2)的第二个数和表示 $B$ 的维数(2×3)的第一个数比较一下,就可以检验这一点。另一方面,颠倒乘积 $BA$ 没有意义,因为 $B$(现在为前一个矩阵)具有 3 列,而 $A$(后一矩阵)仅有一行,因此违背可相乘条件。

一般而言,若 $A$ 为 $m\times n$ 维矩阵,$B$ 为 $p\times q$ 维矩阵,当且仅当 $n=p$ 时,矩阵积 $A\times B$ 有定义。若其有定义,则矩阵积 $A\times B$ 的维数为 $m\times q$——行数与前一矩阵 $A$ 相同,列数与后一矩阵 $B$ 相同。对于在(4.5)中给出的矩阵,$AB$ 的维数为 1×3。

还需要准确定义乘法的计算步骤。为此,我们以(4.5)中的矩阵 $A$ 和 $B$ 为例来加以说明。因为矩阵积 $AB$ 有定义,且预期其维数为 1×3,我们一般可写成(使用符号 $C$ 而非 $c'$ 表示行向量):

$$AB = C = \begin{bmatrix} c_{11} & c_{12} & c_{13} \end{bmatrix}$$

积矩阵 $C$ 中的每一元素以 $c_{ij}$ 表示,它定义为前一矩阵 $A$ 中第 $i$ 行的元素与后一矩阵 $B$ 中的第 $j$ 列的乘积之和。例如,为求 $c_{11}$,我们应取矩阵 $A$ 中的第1行(因 $i=1$),矩阵 $B$ 中的第1列($j=1$)

① 矩阵 $A$ 为行向量时,常以 $a'$ 表示。我们运用符号 $A$ 是为了强调这里解释的乘法法则不仅适用于一向量与一矩阵相乘,也适用于一般矩阵相乘。

(如图4.2上半部分所示)再顺序将这些元素对偶在一处，将每一 61

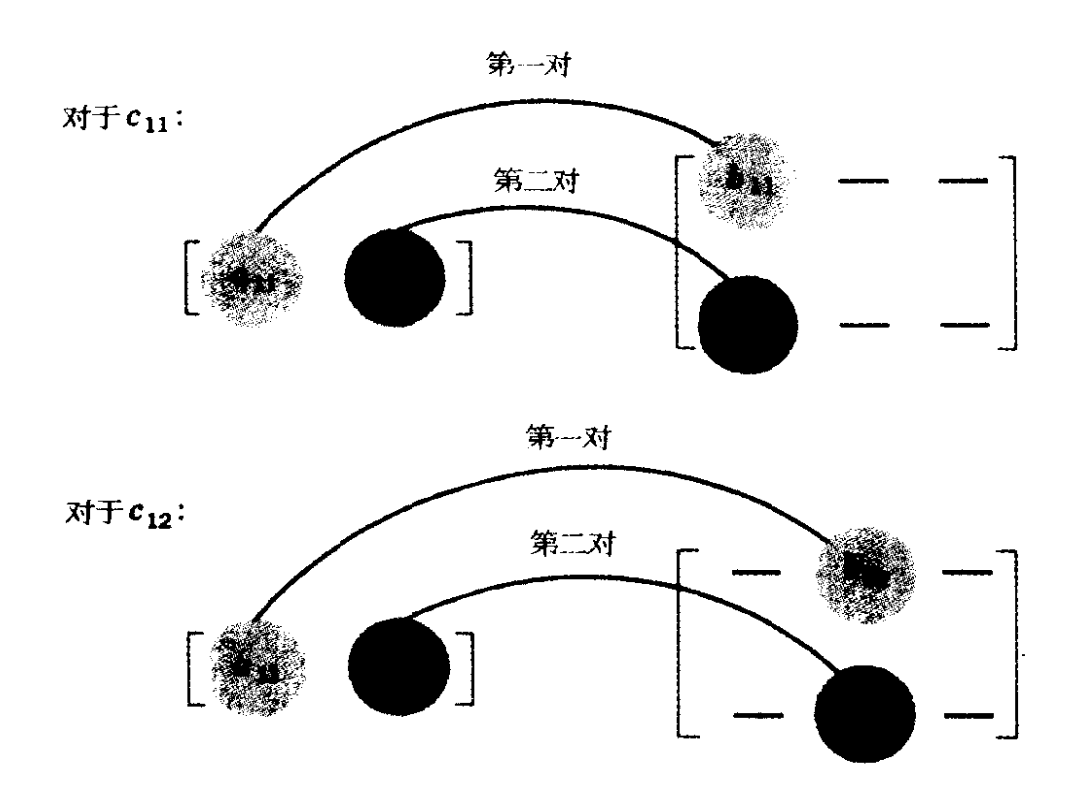



图 4.2

对相乘，然后再对乘积之结果求和，得到

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} \tag{4.6}$$

类似地，对于 $c_{12}$，我们可取 $A$ 中第1行(因 $i=1$)和 $B$ 中第2列(因 $j=2$)，计算指定的乘积和——如图 4.2 的下半部分所示——结果如下：

$$c_{12} = a_{11}b_{12} + a_{12}b_{22} \tag{4.6$'$}$$

以同样方法，我们还有

$$c_{13} = a_{11}b_{13} + a_{12}b_{23} \tag{4.6$''$}$$

在乘积运算能够完成前，正是在此过程中的配对要求必须使得前一矩阵的列维数与后一矩阵的行维数相匹配。

图 4.2 所描述的乘积运算过程还可以用两个向量内积的概念来表达。已知两个向量 $u$ 和 $v$,每个向量具有 $n$ 个元素,如 $(u_1, u_2, \cdots, u_n)$ 和 $(v_1, v_2, \cdots, v_n)$,将其排成两行,或两列,或者一行和一列,其内积以 $u$、$v$ 表示,可以定义为

$$u \cdot v = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

它是对应元素的积的和,因而两个向量的内积为一标量。例如,我们在一次购物后,把采购的 $n$ 种物品数量和其对应顺序的价格作成向量,则其内积给出了全部采购成本。注意,内积的概念不受可相乘条件的约束,因为两个向量排成行或列是无关紧要的。

运用这一概念,我们可以将矩阵积 $C = AB$ 中的元素 $c_{ij}$ 简单地描述为前一矩阵 $A$ 中的第 $i$ 行与后一矩阵中的第 $j$ 列的内积。通过检查图 4.2,我们可以很容易验证这一描述是正确的。

当矩阵 $A$ 和 $B$ 的维数与上述例子不同时,前面描述的乘法法则同样有效。唯一的前提条件是满足可相乘条件。

**例 7** 已知

$$\underset{(2\times2)}{A} = \begin{bmatrix} 3 & 5 \\ 4 & 6 \end{bmatrix} \quad 和 \quad \underset{(2\times2)}{B} = \begin{bmatrix} -1 & 0 \\ 4 & 7 \end{bmatrix}$$

求 $AB$。乘积 $AB$ 显然有定义,其维数为 $2\times2$:

$$AB = \begin{bmatrix} 3(-1) + 5(4) & 3(0) + 5(7) \\ 4(-1) + 6(4) & 4(0) + 6(7) \end{bmatrix} = \begin{bmatrix} 17 & 35 \\ 20 & 42 \end{bmatrix}$$

**例 8** 已知

$$\underset{(3\times2)}{A} = \begin{bmatrix} 1 & 3 \\ 2 & 8 \\ 4 & 0 \end{bmatrix} \quad 和 \quad \underset{(2\times1)}{b} = \begin{bmatrix} 5 \\ 9 \end{bmatrix}$$

求 $Ab$。积矩阵维数应为 $3\times1$,即它是一个列向量

$$Ab = \begin{bmatrix} 1(5)+3(9) \\ 2(5)+8(9) \\ 4(5)+0(9) \end{bmatrix} = \begin{bmatrix} 32 \\ 82 \\ 20 \end{bmatrix}$$

**例 9** 给定

$$\underset{(3\times3)}{A} \begin{bmatrix} 3 & -1 & 2 \\ 1 & 0 & 3 \\ 4 & 0 & 2 \end{bmatrix} \quad 和 \quad \underset{(3\times3)}{B} = \begin{bmatrix} 0 & -\frac{1}{5} & \frac{3}{10} \\ -1 & \frac{1}{5} & \frac{7}{10} \\ 0 & \frac{2}{5} & -\frac{1}{10} \end{bmatrix}$$

求 $AB$。应用同样的运算法则,这次得到一个极特殊的积矩阵

$$AB = \begin{bmatrix} 0+1+0 & -\frac{3}{5}-\frac{1}{5}+\frac{4}{5} & \frac{9}{10}-\frac{7}{10}-\frac{2}{10} \\ 0+0+0 & -\frac{1}{5}+0+\frac{6}{5} & \frac{3}{10}+0-\frac{3}{10} \\ 0+0+0 & -\frac{4}{5}+0+\frac{4}{5} & \frac{12}{10}+0-\frac{2}{10} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

最后一个矩阵——主对角线(从西北到东南的对角线)上的元素为1,其余为0的方阵——给出了一种重要的矩阵类型,即单位矩阵。后面还要对其进行进一步的讨论。

**例 10** 现在我们取如(4.4)所定义的矩阵 $A$ 和向量 $x$,求 $Ax$。积矩阵为一个 $3\times1$ 列向量

63

$$\underset{(3\times3)}{Ax = \begin{bmatrix} 6 & 3 & 1 \\ 1 & 4 & -2 \\ 4 & -1 & 5 \end{bmatrix}} \underset{(3\times1)}{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}} = \underset{(3\times1)}{\begin{bmatrix} 6x_1+3x_2+x_3 \\ x_1+4x_2-2x_3 \\ 4x_1-x_2+5x_3 \end{bmatrix}}$$

重述一遍:尽管右侧的矩阵外形臃肿,但仍是一个向量。因此,当

我们写出 $Ax=d$ 时，我们有：

$$\begin{bmatrix}6x_1+3x_2+x_3\\x_1+4x_2-2x_3\\4x_1-x_2+5x_3\end{bmatrix}=\begin{bmatrix}22\\12\\10\end{bmatrix}$$

根据矩阵相等的定义，上式与(4.3)的整个方程组的表述是相等的。

注意，运用矩阵概念 $Ax=d$，由于可相乘条件，将变量 $x_j$ 排成列向量是必要的，尽管在原方程组中这些变量被排成横的顺序。

**例 11** 包含两个内生变量 $Y$ 和 $C$ 的简单国民收入模型

$$Y=C+I_0+G_0$$

$$C=a+bY$$

可以重排成如(4.1)那样的标准形式

$$Y-C=I_0+G_0$$

$$-bY+C=a$$

因此系数矩阵 $A$，变量向量 $x$，常数向量 $d$ 为：

$$\underset{(2\times2)}{A}=\begin{bmatrix}1&-1\\-b&1\end{bmatrix}\quad\underset{(2\times1)}{x}=\begin{bmatrix}Y\\C\end{bmatrix}\quad\underset{(2\times1)}{d}=\begin{bmatrix}I_0+G_0\\a\end{bmatrix}$$

现在我们验证可以通过方程 $Ax=d$ 来表示该给定方程组。

根据矩阵乘法法则，我们有

$$Ax=\begin{bmatrix}1&-1\\-b&1\end{bmatrix}\begin{bmatrix}Y\\C\end{bmatrix}=\begin{bmatrix}1(Y)+(-1)(C)\\-b(Y)+1(C)\end{bmatrix}=\begin{bmatrix}Y-C\\-bY+C\end{bmatrix}$$

因而矩阵方程 $Ax=d$ 将使我们得到

$$\begin{bmatrix}Y-C\\-bY+C\end{bmatrix}=\begin{bmatrix}I_0+G_0\\a\end{bmatrix}$$

因为矩阵相等意味着对应元素相等，很明显方程 $Ax=d$ 确实代

表了以上面(4.1)形式表示的原方程组。

## 除法问题

尽管矩阵像数一样,可以进行加、减、乘(当然要满足可相乘条件)运算,但却不可能用一个矩阵去除另一个矩阵,即我们不能写出 $A/B$。

对于两个数 $a$ 和 $b$,商 $a/b(b \neq 0)$还可以写出 $ab^{-1}$或 $b^{-1}a$,其中 $b^{-1}$表示 $b$ 的逆或倒数。因为 $ab^{-1} = b^{-1}a$,商式 $a/b$ 既可用于表示 $ab^{-1}$,也可用于表示 $b^{-1}a$。而矩阵的情况则不同,把逆的概念应用于矩阵,我们可以在某些情况下(下面将要讨论)定义矩阵 $B$ 的逆矩阵 $B^{-1}$。但由前面矩阵可乘条件的讨论可知,若 $AB^{-1}$有定义,但不能保证 $B^{-1}A$ 也有定义。即使 $AB^{-1}$和 $B^{-1}A$ 均有定义,但二者未必表示同样的积。所以 $A/B$ 不能做到明确无误地应用,因而应避免使用这种表达方式。若逆矩阵 $B^{-1}$确实存在,且所讨论的矩阵积有定义,必须指明所指的是 $AB^{-1}$还是 $B^{-1}A$。逆矩阵后面还要进一步讨论。

## 对"∑"符号的题外探讨

运用下标符号不仅有助于标明参数和变量的位置,而且有利于灵活地简化加和项的表示,正如在矩阵积运算过程中出现的那样。

运用希腊字母$\sum$(sigma,表示"和")表示和的简写形式。例如,要表示 $x_1, x_2, x_3$ 的和,可以写成

$$x_1 + x_2 + x_3 = \sum_{j=1}^{3} x_j$$

此式读作:“当 $j$ 取 1 至 3 时,$x_j$ 的和”。符号 $j$,称作求和指数,仅取整数值。$x_j$ 表示被加数(即将要被加的数),它实际上是 $j$ 的函数。除了字母 $j$ 外,求和指数也常以 $i$ 或 $k$ 表示,比如

$$\sum_{i=3}^{7} x_i = x_3 + x_4 + x_5 + x_6 + x_7$$

$$\sum_{k=0}^{n} x_k = x_0 + x_1 + \cdots + x_n$$

65

$\sum$符号的应用可极易推广到以下情况:$x$ 项前具有系数或和中各项具有整数幂。例如,我们可以写:

$$\sum_{j=1}^{3} ax_j = ax_1 + ax_2 + ax_3 = a(x_1 + x_2 + x_3) = a\sum_{j=1}^{3} x_j$$

$$\sum_{j=1}^{3} a_j x_j = a_1 x_1 + a_2 x_2 + a_3 x_3$$

$$\sum_{i=0}^{n} a_i x^i = a_0 x^0 + a_1 x^1 + a_2 x^2 + \cdots + a_n x^n$$

$$= a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

特别地,最后一个例子说明,表达式 $\sum_{i=0}^{n} x^i$ 实际上可以用作一般多项式(2.4)的简写形式。

顺便提一句,当所讨论问题的求和范围在上下文中非常明确时,符号$\sum$可以单独使用,无需附上求和指数(如 $\sum x_i$ ),或者在下面仅标上一个指数字母(如 $\sum_i x_i$ )。

我们将简写符号$\sum$应用于矩阵积。在(4.6)、(4.6′)及(4.6″)中,积矩阵 $C = AB$ 的每一元素被定义为各项和。现在可将其改写成如下形式:

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} = \sum_{k=1}^{2} a_{1k}b_{k1}$$

$$c_{12} = a_{11}b_{12} + a_{12}b_{22} = \sum_{k=1}^{2} a_{1k}b_{k2}$$

$$c_{13} = a_{11}b_{13} + a_{12}b_{23} = \sum_{k=1}^{2} a_{1k}b_{k3}$$

在每一种情况下，$c_{1j}$的第 1 个下标在$\sum$表达式中反映在 $a_{1k}$的第 1 个下标上，$c_{1j}$的第 2 个下标在$\sum$表达式中反映在 $b_{kj}$的第 2 个下标上。另外，指标 $k$ 是一个“哑”下标，它用于表示哪一特定的元素偶相乘，但并未在符号 $c_{1j}$中出现。

将这种简写方式拓展到 $m \times n$ 维矩阵 $A = [a_{ik}]$和 $n \times p$ 维矩阵 $B = [b_{kj}]$相乘的情况。现在我们可以将 $m \times p$ 积矩阵 $AB = C = [c_{ij}]$中的每一元素写成

$$c_{11} = \sum_{k=1}^{n} a_{1k}b_{k1} \qquad c_{12} = \sum_{k=1}^{n} a_{1k}b_{k2} \cdots$$

或者更一般地

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} \qquad \begin{pmatrix} i = 1,2,\cdots,m \\ j = 1,2,\cdots,p \end{pmatrix}$$

66

最后一个方程还为前面所定义的矩阵乘法法则给出了另一种表达方式。

## 练习 4.2

---

1　已知 $A = \begin{bmatrix} 4 & -1 \\ 6 & 9 \end{bmatrix}$，$B = \begin{bmatrix} 0 & 3 \\ 3 & -2 \end{bmatrix}$，和 $C = \begin{bmatrix} 8 & 3 \\ 6 & 1 \end{bmatrix}$，求：

(a) $A+B$　　(b) $C-A$　　(c) $3A$　　(d) $4B+2C$

2　已知 $A=\begin{bmatrix}2 & 8\\3 & 0\\5 & 1\end{bmatrix}$, $B=\begin{bmatrix}2 & 0\\3 & 8\end{bmatrix}$, 和 $C=\begin{bmatrix}7 & 2\\6 & 3\end{bmatrix}$:

(a) $AB$ 有定义吗? 计算 $AB$。可以计算 $BA$ 吗? 为什么?

(b) $BC$ 是否有定义? 计算 $BC$。$CB$ 是否有定义? 若有,计算之。$BC=CB$ 成立吗?

3　在例 9 给出的矩阵的基础上,判断 $BA$ 是否有定义? 若有,计算其积。在此例中,我们是否有 $AB=BA$?

4　求下列积矩阵(在每一积矩阵下面标上维数):

(a) $\begin{bmatrix}0 & 1 & 0\\3 & 0 & 4\\2 & 3 & 0\end{bmatrix}\begin{bmatrix}8 & 0\\0 & 1\\3 & 5\end{bmatrix}$　　(c) $\begin{bmatrix}3 & 2 & 0\\4 & 2 & -7\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}$

(b) $\begin{bmatrix}6 & 5 & 1\\3 & 0 & 4\end{bmatrix}\begin{bmatrix}4 & -1\\5 & 2\\0 & 1\end{bmatrix}$　　(d) $\begin{bmatrix}a & b & c\end{bmatrix}\begin{bmatrix}7 & 0\\0 & 2\\1 & 4\end{bmatrix}$

5　展开下列和表达式:

(a) $\sum_{i=2}^{5} x_i$　　(c) $\sum_{i=1}^{4} bx_i$　　(e) $\sum_{i=0}^{3}(x+i)^2$

(b) $\sum_{i=5}^{8} a_i x_i$　　(d) $\sum_{i=1}^{n} a_i x^{i-1}$

6　用$\sum$符号改写下列各式:

(a) $x_1(x_1-1)+2x_2(x_2-1)+3x_3(x_3-1)$

(b) $a_2(x_3+2)+a_3(x_4+3)+a_4(x_5+4)$

(c) $\frac{1}{x}+\frac{1}{x^2}+\cdots+\frac{1}{x^n}\ (x\neq 0)$

(d) $1+\frac{1}{x}+\frac{1}{x^2}+\cdots+\frac{1}{x^n}\ (x\neq 0)$

7　证明下列各式成立:

(a) $\left(\sum_{i=0}^{n} x_i\right)+x_{n+1}=\sum_{i=0}^{n+1} x_i$

(b) $\sum_{j=1}^{n} ab_j y_j = a\sum_{j=1}^{n} b_j y_j$

67

(c) $\sum_{j=1}^{n}(x_j + y_j) = \sum_{j=1}^{n} x_j + \sum_{j=1}^{n} y_j$

---

## 4.3 对向量运算的注释

68

在前面,向量被视为矩阵的一种特殊类型。因此,它们适用于所讨论的所有代数运算。然而,由于向量维数的特殊性,对其作额外的讨论是有益的。

**向量乘法**

一个 $m\times 1$ 列向量 $u$ 与一 $1\times n$ 行向量相乘,得到 $m\times n$ 维积矩阵 $uv'$。

**例 1** 已知 $u=\begin{bmatrix}3\\2\end{bmatrix}$和 $v'=[1\ 4\ 5]$,可以得到

$$uv' = \begin{bmatrix}3(1) & 3(4) & 3(5)\\ 2(1) & 2(4) & 2(5)\end{bmatrix} = \begin{bmatrix}3 & 12 & 15\\ 2 & 8 & 10\end{bmatrix}$$

因为矩阵 $u$ 中的每一行仅有一个元素,$v'$中每一列也仅有一个元素,结果 $uv'$中的每一元素仅是一个单积而非积的和。尽管我们仅用两个向量相乘,但积矩阵 $uv'$却是一个 $2\times 3$ 维矩阵。

另一方面,已知一个 $1\times n$ 行向量 $u'$和一个 $n\times 1$ 列向量 $v$,积 $u'v$ 将是一个 $1\times 1$ 维矩阵。

**例 2** 已知 $u'=[3\ 4]$和 $v=\begin{bmatrix}9\\7\end{bmatrix}$,我们有

$$u'v = [3(9)+4(7)] = [55]$$

$u'v$ 尽管只有一个元素,但正如上面所写出的那样,它仍是一个矩

阵。然而,$1\times1$ 矩阵就加法和乘法而言,完全类似于标量:$[4]+[8]=[12]$恰如 $4+8=12$;$[3][7]=[21]$恰如 $3(7)=21$。此外,$1\times1$矩阵并不具有标量所不具备的性质。事实上,所有标量集合与元素为标量的所有 $1\times1$ 矩阵集合之间存在一一对应的关系。因此,我们可以将 $u'v$ 定义为对应于 $1\times1$ 积矩阵的标量。对于上面的例子,我们可相应地写成 $u'v=55$。这样的积称作为标量集。[①] 但要记住,尽管 $1\times1$ 矩阵可以看作标量,但在矩阵进一步计算过程中,不能随意以 $1\times1$ 矩阵代替标量,除非其满足可相乘条件。

**例3** 已知行向量 $u'=[3\ 6\ 9]$,求 $u'u$。因为 $u$ 仅是一个将 $u'$的元素纵排的列向量,我们有

$$u'u=[3\ 6\ 9]\begin{bmatrix}3\\6\\9\end{bmatrix}=(3)^2+(6)^2+(9)^2$$

这里我们省略了右侧 $1\times1$ 积矩阵中的括号[ ]。注意,积 $u'u$ 给出了 $u$ 的元素的平方和。

一般而言,若 $u'=[u_1,u_2,\ \cdots,\ u_n]$,则 $u'u$ 将是元素 $u_j$ 的平方和(一个标量):

$$u'u=u_1{}^2+u_2{}^2+\cdots+u_n{}^2=\sum_{j=1}^{n}u_j^2$$

若我们计算内积 $u\cdot u$(或 $u'\cdot u'$),当然会得到完全相同的结果。

---

① 标量积的概念类似于两个元素数量相同向量的内积概念,后者也得到一个标量。但读者回忆一下,内积是不要求可相乘条件的,所以可写成 $u\cdot v$。另一方面,在标量积情况下(两个向量符号之间无小圆点),我们仅能将其表示为一个行向量被列向量乘,行向量居前。

总之，区分 $uv'$（大于 1×1 的矩阵）与 $u'v$（一个 1×1 矩阵或 69
标量）的含义是重要的。特别要注意，标量积一定有行向量居前，列向量居后，否则积矩阵不可能是 1×1 矩阵。

## 向量运算的几何解释

前面已提到，具有 $n$ 个元素的行向量或列向量（以后称作 $n$－向量）可视作 $n$ 元数组，并还可视作 $n$ 维空间中的（以后称 $n$－空间）一个点。下面我们来详细讨论这一思想。在图 4.3$a$ 中，一个点（3,2）被画在二维空间中并标以 $u$。这是向量 $u=\begin{bmatrix}3\\2\end{bmatrix}$或向量 $u'=[3\ 2]$的几何表示。在这里这两个向量代表同一个相同的有序偶。若从原点（0,0）向点 $u$ 画一箭头（方向线段），则它规定了从原点到 u 点的唯一一条直线。因为对于每一点有唯一的方向线段，所以在图形上，我们既可以把向量 $u$ 以点（3,2）表示，也可以相应的方向线段表示。这样从原点（0,0）发出的方向线段，像表的指针，有确定的长度和方向，被称作矢经。

由这个对向量的新的解释，可以给出下列运算的几何含义：（$a$）向量的标量积；（$b$）向量的加减法，以及更一般地，（$c$）所谓向量的“线性组合”。

首先，若我们在图 4.3$a$ 中画出向量$\begin{bmatrix}6\\4\end{bmatrix}=2u$，相应的方向线段将覆盖原有方向线段并为其两倍长。事实上，向量 $u$ 被任意标量相乘将产生一重叠的方向线段，但除非 $k=1$，否则箭头的位置 70
将重新确定。若标量乘子 $k>1$，则方向线段会延长（比例扩大）；若 $0<k<1$，则方向线段会缩短（比例缩小）。若 $k=0$，则方向线

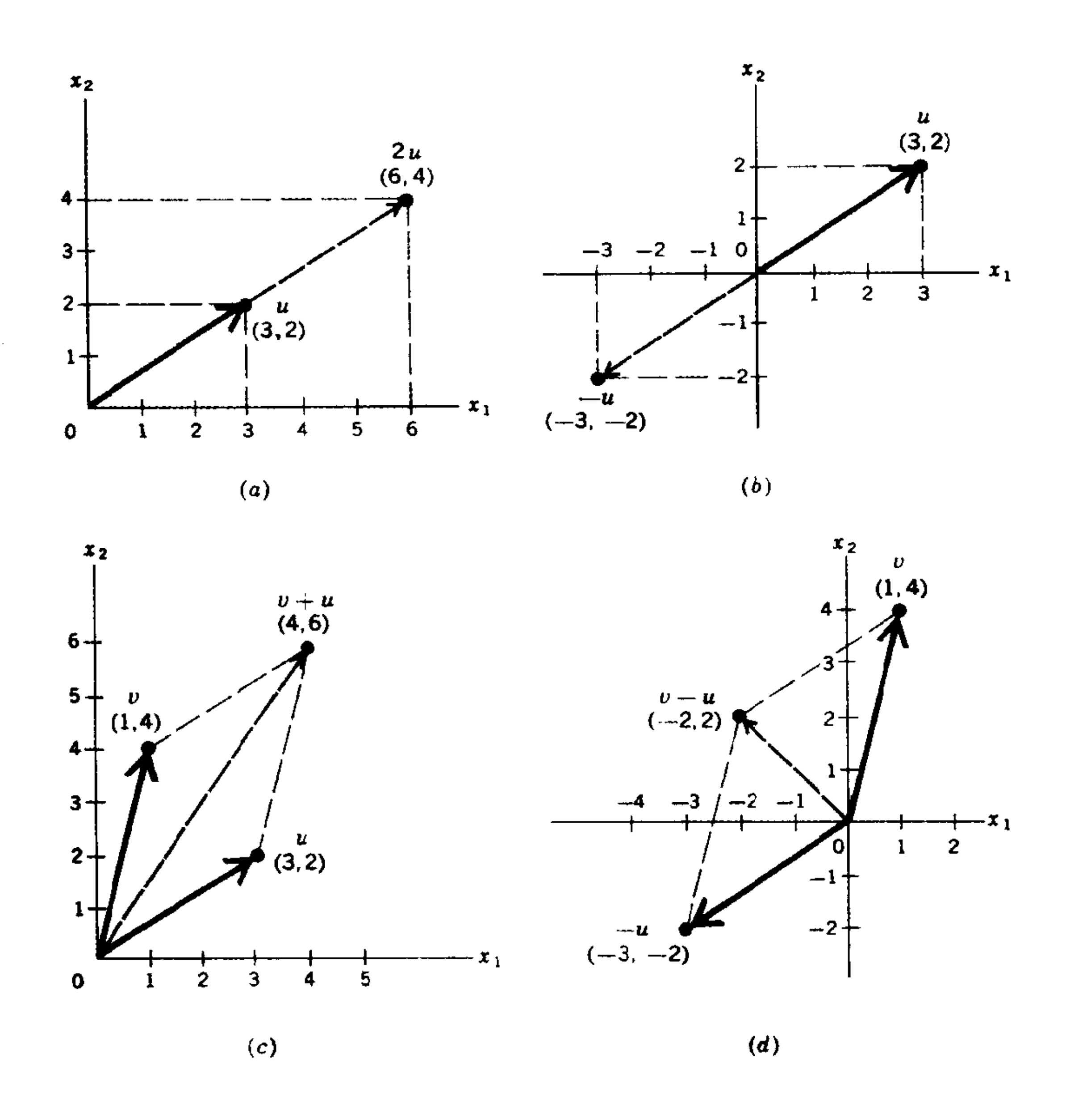



图 4.3

段会缩回到原点——它表示零向量，$\begin{bmatrix}0\\0\end{bmatrix}$。负的标量乘子甚至会倒转方向线段的方向。例如，向量 $u$ 被 $-1$ 乘，我们得到 $-u=\begin{bmatrix}-3\\-2\end{bmatrix}$，将其在图 4.3$b$ 中画出，它是一条与 $u$ 长度相同，但方向完全相反的方向线段。

其次，我们考察两个向量 $v=\begin{bmatrix}1\\4\end{bmatrix}$，$u=\begin{bmatrix}3\\2\end{bmatrix}$的加法。其和 $v+$

$u=\begin{bmatrix}4\\6\end{bmatrix}$可直接在图 4.3$c$ 中以虚方向线段画出。若以向量 $u$ 和 $v$（实线）为两边作平行四边形，此平行四边形的对角线恰是表示向量和 $v+u$ 的方向线段。一般而言，向量和在几何上可由平行四边形得到。进而，这种方法还可以使我们得到向量差 $v-u$。因为后者等价于 $v$ 与 $(-1)u$ 之和。在图 4.3$d$ 中，我们首先由图形 $b$ 和 $c$ 分别重新画出了向量 $v$ 和 $-u$，然后再画出平行四边形，其对角线便表示向量差 $v-u$。

将上述结论作简单的推广便可以对向量的线性组合（即线性和或差）作出几何解释。考察下面的简单例子，

$$3v+2u=3\begin{bmatrix}1\\4\end{bmatrix}+2\begin{bmatrix}3\\2\end{bmatrix}=\begin{bmatrix}9\\16\end{bmatrix}$$

这个运算的标量乘法分别使两向量 $v$ 和 $u$ 的箭头重新定位，加法则要求作出平行四边形。除了这些基本的图解运算外，向量的线性组合并无新内容；即使线性组合涉及更多的项也同样如此。比如，

$$\sum_{i=1}^{n}k_iv_i=k_1v_1+k_2v_2+\cdots+k_nv_n$$

其中 $k_i$ 为标量集合，而 $v_i$ 则表示向量集合。要想求得此和，应首先将前两项相加，然后将其和再与第三项相加，这样不断相加，就可以求出各项和。

## 线性相关

当且仅当一组向量 $v_1$，$v_2$，$\cdots$，$v_n$ 中任一向量可以表示为其它向量的线性组合时，我们称这组向量为线性相关的，否则便是线性无关的。

71 **例4** 三个向量 $v_1=\begin{bmatrix}2\\7\end{bmatrix}$, $v_2=\begin{bmatrix}1\\8\end{bmatrix}$以及 $v_3=\begin{bmatrix}4\\5\end{bmatrix}$是线性相关的,因为 $v_3$ 是 $v_1$ 与 $v_2$ 的线性组合:

$$3v_1-2v_2=\begin{bmatrix}6\\21\end{bmatrix}-\begin{bmatrix}2\\16\end{bmatrix}=\begin{bmatrix}4\\5\end{bmatrix}=v_3$$

注意,上一个方程也可以表示成

$$3v_1-2v_2-v_3=0$$

其中 $0\equiv\begin{bmatrix}0\\0\end{bmatrix}$,表示零向量。

**例5** 两个行向量 $v_1'=[5\ 12]$和 $v_2'=[10\ 24]$是线性相关的,因为

$$2v_1'=2[5\ 12]=[10\ 24]=v_2'$$

事实上,一个向量是另一个向量的倍数是向量线性组合的最简单的情况。再提醒一下,上一个方程可以等价地写成

$$2v_1'-v_2'=0'$$

其中 $0'$表示 0 行向量[0 0]。

运用零向量的概念,线性相关可以重新定义如下:一组 $m$-向量 $v_1, v_2, \cdots, v_n$ 为线性相关的当且仅当存在一组标量 $k_1$, $k_2$, $\cdots$, $k_n$(不全为 0)使得

$$\sum_{i=1}^{n}k_iv_i=\underset{(m\times1)}{0}$$

另一方面,若对所有的 $i$,仅当 $k_i=0$ 时,上述方程才成立,则这些向量是线性无关的。

线性相关的概念也可以很容易给出几何解释。两个向量 $u$ 和 $2u$,——一个为另一个的倍数——明显是线性相关的。在图

4.3*a* 中，从图形上看，其箭头位于同一条直线上。图 4.3*b* 中的两个相关向量 $u$ 和 $-u$ 也同样如此。相反，图 4.3*c* 中的两个向量 $u$ 和 $v$ 是**线性无关**的，因为不可能将其中一个表示成另一个的倍数。从几何上看，其箭头并不位于同一条直线上。

在二维空间中考察两个以上的向量，可以得出这样一个重要结论：在 2－空间中，一旦我们找到两个线性无关的向量（比如 $u$ 和 $v$），该空间中所有其它向量均可表示为这两个向量（$u$ 和 $v$）的线性组合。在图 4.3*c* 和 *d* 中，我们已经说明了如何求出两个简单的线性组合 $v+u$ 和 $v-u$。进而，通过延长，缩短，旋转已知向量 72 $u$ 和 $v$，然后将其构造成各种平行四边形，可以得到无数新的向量，将所有的 2－向量包罗无遗。因此，任意三个或更多的 2－向量（2－空间中三个或三个以上的向量）必定是线性相关的。其中的两个向量可能是线性无关的，但第三个必定是前两个的线性组合。

## 向量空间

两个线性无关的向量 $u$ 和 $v$ 的各种线性组合生成的所有 2－向量的全体称为二维**向量空间**。因为我们现在仅考察具有实值元素的向量，而该向量空间恰好是 $R^2$，即我们一直所说的 2－空间。2－空间不能由单个 2－向量生成，因为单个 2－向量的“线性组合”只能生成位于一条直线上的向量集合。2－空间的生成也无需两个以上线性无关的 2－向量——无论如何，不可能找到两个以上线性无关的 2－向量。

我们说，两个线性无关的向量 $u$ 和 $v$ **生成**了 2－空间，还可以说二者构成了 2－空间的某个**基**。注意，我们说二者构成 2－空间

的“某个基”，而不说二者构成 2－空间的“基”，是因为任何一对 2－向量，只要它们是线性无关的，都可以构成 2－空间的一个基。特别地，考察两个单位向量[1 0]和[0 1]。前者可画成一个位于横轴上的一个箭头（方向线段），后者可画成一个位于纵轴的方向线段。因为它们是线性无关，所以可作为 2－空间的一个基，而且我们通常确实认为 2－空间由两个轴生成，而两轴不过是两个单位向量的延伸。

类似地，三维向量空间是 3－向量的全体，它由三个线性无关的 3－向量生成。作为一个说明，考察三个单位向量集

$$e_1 \equiv \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad e_2 \equiv \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad e_3 \equiv \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{4.7}$$

其中 $e_i$ 为第 $i$ 个元素为 1，其余元素为零的向量。这三个向量显然是线性无关的。事实上，其方向线段位于图 4.4 三维空间的三个轴上。因而他们生成了3－空间。这意味着整个 3－空间（在我们的分析框架中为 $R^3$）可由这些单位向量生成。例如，向量 $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ 可视为 $e_1+2e_2+2e_3$ 的线性组合。在几何上，我们可以首先用平行四边形法将 $e_1$ 和 $2e_2$ 在图 4.4 中相加，在 $x_1x_2$ 平面上得到以点(1,2,0)表示的向量，然后再将此向量与 $2e_3$ 相加——在阴影垂直平面上作平行四边形——在点(1,2,2)得到了所希望的最终结果。

向 $n$－空间推广也是显然的。$n$－空间可以定义为 $n$ 向量的全体。尽管难以画出图形，我们仍可想象 $n$－空间由全部线性无

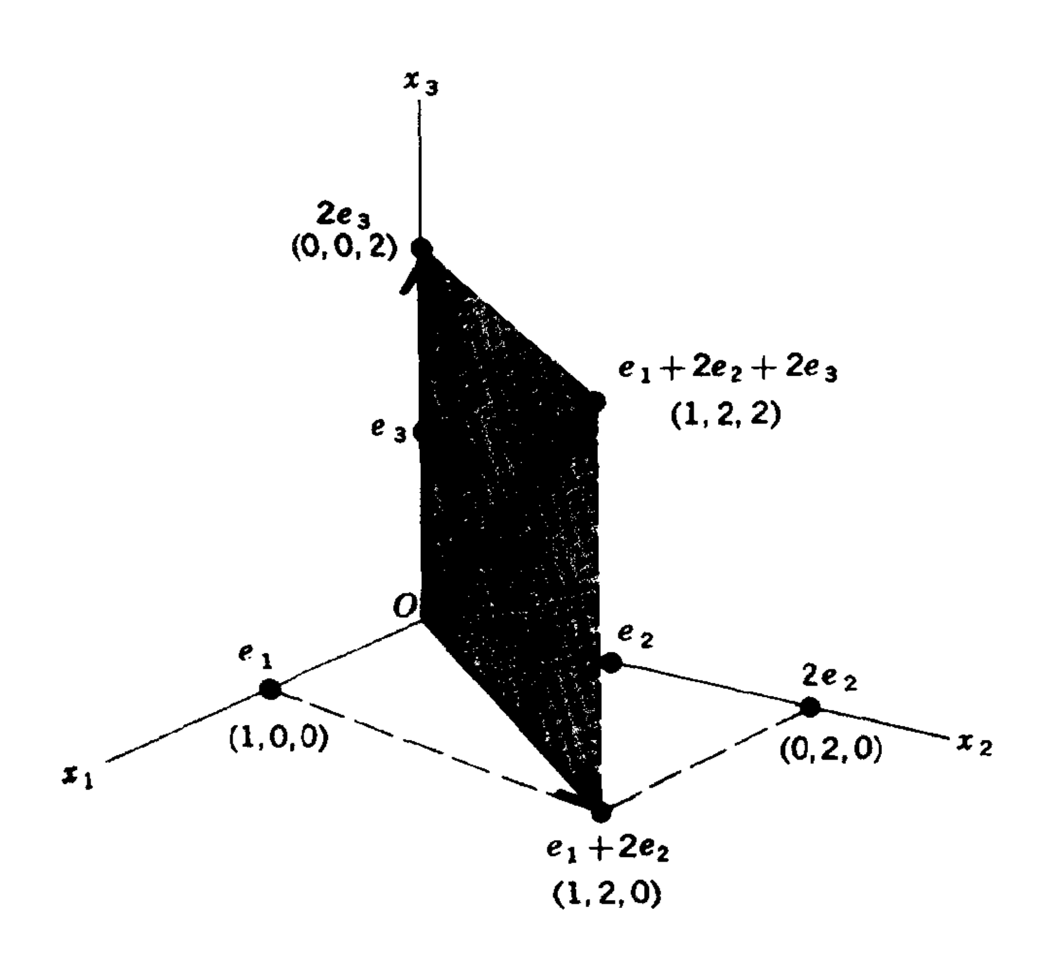



图 4.4

关的 $n$($n$－元素)个单位向量的全体所生成。每一个 $n$－向量，作为一个有序 $n$ 元数组，代表 $n$－空间中的一个点，或者由原点 73 (即 $n$－元素零向量)延伸至该点的方向线段。事实上，任意给定的 $n$ 个线性无关的 $n$－向量的集合，都能生成整个 $n$－空间。因为在我们的讨论中，$n$－向量的每一元素限定为一个实数，该 $n$－空间实际上是 $R^n$。

上面所指的 $n$－空间有时更特别地称作欧几里得 $n$－空间(以欧几里得名字命名)。为对后一概念进行解释，我们首先简要地介绍一下两个向量点之间的距离的概念。对于给定空间中任意一对向量点 $u$ 和 $v$，$u$ 到 $v$ 的距离是某个具有下述性质的实值函数

$$d = d(u, v)$$

(1) 当 $u$ 和 $v$ 重合时，其距离为零；(2) 当两个点不重合时，从 $u$ 和 $v$ 的距离和从 $v$ 到 $u$ 的距离由同一正实数表示；(3) $u$ 到 $v$ 的距离永不会长于 $u$ 到 $w$（一个与 $u$ 和 $v$ 不同的点）的距离与 $w$ 到 $v$ 的距离之和。以符号表示，

$$d(u,v) = 0 \qquad (\text{因 } u = v)$$

$$d(u,v) = d(v,u) > 0 \qquad (\text{因 } u \neq v)$$

$$d(u,v) \leqslant d(u,w) + d(w,v) \qquad (\text{因 } w \neq u, v)$$

74 最后一个性质被称作三角形不等式，因为这个点 $u, v, w$ 通常可定义一个三角形。

当一个向量空间定义了一个满足上述三个性质的距离函数时，则称其为一个度量空间。但要注意，前面只是对距离 $d(u,v)$ 作了一般意义的讨论。当 $d$ 函数取各种特定形式时，能够产生各种度量空间。所谓“欧几里得空间”是度量空间中的一种特殊形式，它具有如下定义的距离函数：令点 $u$ 为 $n$ 元数组 $(a_1, a_2, \cdots, a_n)$，点 $v$ 为 $n$ 元数组 $(b_1, b_2, \cdots, b_n)$，则欧几里得距离函数为

$$d(u,v) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}$$

这里取正平方根。容易验证，这个特殊距离函数满足上述三个性质。将其应用于图 4.3$a$ 中的二维空间，可求得点 $(6,4)$ 与 $(3,2)$ 间的距离为

$$\sqrt{(6-3)^2 + (4-2)^2} = \sqrt{3^3 + 2^2} = \sqrt{13}$$

这一结果与毕达哥拉斯定理是一致的。此定理表明，直角三角形的斜边等于另外两边平方和的（正）平方根。因为若我们将点 $(6,4)$ 和 $(3,2)$ 看作 $u$ 和 $v$，并在 $(6,2)$ 画出一个新的点 $w$，则我们

得到一个直角三角形，其横直角边和纵直角边分别为 3 和 2，斜边长度（$u$ 和 $v$ 的距离）等于 $\sqrt{3^2+2^2}=\sqrt{13}$。

欧几里得距离函数也可以用两个向量的标量积的平方根来表示。因为 $u$ 和 $v$ 表示两个 $n$ 元数组 $(a_1, a_2, \cdots, a_n)$ 和 $(b_1, b_2, \cdots, b_n)$，所以我们可以写出列向量 $u-v$，其元素为 $a_1-b_1$，$a_2-b_2$，$\cdots$，$a_n-b_n$。欧几里得距离函数中平方根号下的数当然就等于这 $n$ 个元素的平方和，参照上面例 3，可将其写成标量积 $(u-v)'(u-v)$，因此我们有

$$d(u,v)=\sqrt{(u-v)'(u-v)}$$

## 练习 4.3

---

1　已知 $u'=[5\ 2\ 3]$，$v'=[3\ 1\ 9]$，$w'=[7\ 5\ 8]$，和 $x'=[x_1\ x_2\ x_3]$，写出列向量 $u$，$v$，$w$ 和 $x$，并求

(a) $uv'$　(c) $xx'$　(e) $u'v$　(g) $u'u$

(b) $uw'$　(d) $v'u$　(f) $w'x$　(h) $x'x$

2　已知 $w=\begin{bmatrix}3\\2\\16\end{bmatrix}$，$x=\begin{bmatrix}x_1\\x_2\end{bmatrix}$，$y=\begin{bmatrix}y_1\\y_2\end{bmatrix}$，和 $z=\begin{bmatrix}z_1\\z_2\end{bmatrix}$：

75

(a) 下列乘积哪些有定义？

$$w'x,\ x'y',\ xy',\ y'y,\ zz',\ yw',\ x\cdot y?$$

(b) 求出有定义的所有积。

3　购买 $n$ 种商品，数量为 $Q_1$，$Q_2$，$\cdots$，$Q_n$，价格 $P_1, P_2, \cdots, P_n$，如何以 (a) 符号 $\sum$ 和 (b) 向量概念表示全部采购成本？

4　给定两个非零向量 $w_1$ 和 $w_2$，二者间形成的角 $Q$ $(0^\circ \leqslant Q \leqslant 180^\circ)$ 以下列方式与标量积 $w_1'w_2(=w_2'w_1)$ 相联系：

$$Q\text{ 是一}\begin{Bmatrix}\text{锐角}\\\text{直角}\\\text{钝角}\end{Bmatrix}\text{当且仅当 } w_1'w\begin{Bmatrix}>\\=\\<\end{Bmatrix}0$$

通过计算下面每一对向量的标量积验证这一结论：

(a) $w_1=\begin{bmatrix}3\\2\end{bmatrix}, w_2=\begin{bmatrix}1\\4\end{bmatrix}$　　(d) $w_1=\begin{bmatrix}1\\0\\0\end{bmatrix}, w_2=\begin{bmatrix}0\\2\\0\end{bmatrix}$

(b) $w_1=\begin{bmatrix}1\\4\end{bmatrix}, w_2=\begin{bmatrix}-3\\-2\end{bmatrix}$　　(e) $w_1=\begin{bmatrix}1\\2\\2\end{bmatrix}, w_2=\begin{bmatrix}1\\2\\0\end{bmatrix}$

(c) $w_1=\begin{bmatrix}3\\2\end{bmatrix}, w_2=\begin{bmatrix}-3\\-2\end{bmatrix}$

5　已知 $u=\begin{bmatrix}5\\1\end{bmatrix}$，$v=\begin{bmatrix}0\\3\end{bmatrix}$，求下列向量的图形解：

(a) $2v$　　(c) $u-v$　　(e) $2u+3v$

(b) $u+v$　　(d) $v-u$　　(f) $4u-2v$

6　因为 3－空间由(4.7)定义的三个单位向量所生成，任意其它 3－向量均可表示成 $e_1, e_2, e_3$ 的线性组合。证明下列 3－向量可表示为 $e_1$，$e_2$，$e_3$ 的线性组合：

(a) $\begin{bmatrix}4\\7\\0\end{bmatrix}$　　(b) $\begin{bmatrix}15\\-2\\1\end{bmatrix}$　　(c) $\begin{bmatrix}-1\\3\\9\end{bmatrix}$　　(d) $\begin{bmatrix}2\\0\\8\end{bmatrix}$

7　在三维欧几里得空间中，下列两点之间的距离为多少？

(a) (3,2,8)和(0,－1,5)　　(b) (9,0,4)和(2,0,－4)

8　三角形不等式以弱不等式符号≤表示，而非以严格不等号<表示。在何情况下弱不等式中“＝”部分能够应用。

9　分别用

(a) 标量　(b) 标量积　(c) 内积

表示欧几里得 $n$－空间向量半径的长度(即从原点到 $v$ 点的距离)。

## 4.4 交换律、结合律、分配律



在普通标量代数中,加法和乘法运算满足下列交换律、结合律、分配律:

加法的交换律　　$a+b=b+a$

乘法的交换律　　$ab=ba$

加法的结合律　$(a+b)+c=a+(b+c)$

乘法的结合律　　$(ab)c=a(bc)$

分配律　　$a(b+c)=ab+ac$

在类似的具有同样名称的定律应用于并集和交集的讨论时,也涉及到上述定律,这些定律的大多数均可应用于矩阵运算——乘法的交换律除外。

### 矩阵加法

矩阵加法不仅可运用结合律,也可运用交换律。从下述事实可知:矩阵加法仅要求两个对应元素相加,但相加的顺序无关紧要。对此,顺便说一句,减法运算 $A-B$ 可简单视为加法运算 $A+(-B)$。这样就没有必要单独讨论减法运算了。

**交换律**　　$A+B=B+A$

证明　$A+B=[a_{ij}]+[b_{ij}]=[a_{ij}+b_{ij}]=[b_{ij}+a_{ij}]=B+A$

**例 1**　已知 $A=\begin{bmatrix}3 & 1\\0 & 2\end{bmatrix}$ 和 $B=\begin{bmatrix}6 & 2\\3 & 4\end{bmatrix}$,求得

$$A+B=B+A=\begin{bmatrix}9 & 3\\3 & 6\end{bmatrix}$$

**结合律**：　　$(A+B)+C=A+(B+C)$

证明：$(A+B)+Cp=[a_{ij}+b_{ij}]+[c_{ij}]=[a_{ij}+b_{ij}+c_{ij}]$

$$=[a_{ij}]+[b_{ij}+c_{ij}]=A+(B+C)$$

**例 2**　已知 $v_1=\begin{bmatrix}3\\4\end{bmatrix}$，$v_2=\begin{bmatrix}9\\1\end{bmatrix}$，和 $v_3=\begin{bmatrix}2\\5\end{bmatrix}$，求得

$$(v_1+v_2)-v_3=\begin{bmatrix}12\\5\end{bmatrix}-\begin{bmatrix}2\\5\end{bmatrix}=\begin{bmatrix}10\\0\end{bmatrix}$$

77 它也等于

$$v_1+(v_2-v_3)=\begin{bmatrix}3\\4\end{bmatrix}+\begin{bmatrix}7\\-4\end{bmatrix}=\begin{bmatrix}10\\0\end{bmatrix}$$

结合律应用于向量线性组合 $k_1v_1+\cdots+k_nv_n$，使我们得以选择任意两项首先相加(减)，不必遵从线性组合中 $n$ 项的排列顺序。

### 矩阵乘法

矩阵乘法不服从交换律，即

$$AB\neq BA$$

如前面所解释的那样，即使 $AB$ 有定义，$BA$ 也未必有；即使 $AB$，$BA$ 均有定义，一般的规律仍是 $AB\neq BA$。

**例 3**　　令 $A=\begin{bmatrix}1&2\\3&4\end{bmatrix}$和 $B=\begin{bmatrix}0&-1\\6&7\end{bmatrix}$，则

$$AB=\begin{bmatrix}1(0)+2(6)&1(-1)+2(7)\\3(0)+4(6)&3(-1)+4(7)\end{bmatrix}=\begin{bmatrix}12&13\\24&25\end{bmatrix}$$

但 $BA=\begin{bmatrix}0(1)-1(3)&0(2)-1(4)\\6(1)+7(3)&6(2)+7(4)\end{bmatrix}=\begin{bmatrix}-3&-4\\27&40\end{bmatrix}$

**例 4**　令 $u'$为 $1\times3$(行向量)，则相应的列向量必定为 $3\times1$。积 $u'u$ 必定为 $1\times1$ 矩阵，但积 $uu'$将是 $3\times3$ 矩阵。因此很明显，

$u'u \neq uu'$。

鉴于一般规律 $AB \neq BA$ 成立，我们常用术语左乘和右乘来规定乘积的顺序。在积 $AB$ 中，我们说矩阵 $B$ 被矩阵 $A$ 左乘，矩阵 $A$ 被 $B$ 右乘。

然而，规律 $AB \neq BA$ 确实存在令人感兴趣的例外。一种情况是 $A$ 为方阵，$B$ 为单位矩阵；另一种情况是 $A$ 是 $B$ 的逆矩阵，即 $A = B^{-1}$。这两种例外我们后面还要讨论。这里我们还要强调矩阵的标量积确实服从交换律。因此，

$$kA = Ak$$

若 $k$ 为一标量。

**结合律** $$(AB)C = A(BC) = ABC$$

在形成乘积 $ABC$ 时，每一相邻的矩阵偶必须自然满足可相乘条件。若 $A$ 为 $m \times n$ 矩阵，$C$ 为 $P \times q$ 矩阵，则可相乘条件要求矩阵 $B$ 为 $n \times p$： 78

$$\underset{(m\times n)}{A} \quad \underset{(n\times p)}{B} \quad \underset{(p\times q)}{C}$$

注意，$n$ 和 $p$ 在维数标示中出现两次。若可相乘条件得以满足，结合律表明，任何相邻的一对矩阵必然可首先相乘，只要假设乘积准确地插在初始矩阵偶的恰当位置上。

**例 5** 如果 $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ 和 $A = \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}$，那么

$$x'Ax = x'(Ax) = [x_1 \ x_2]\begin{bmatrix} a_{11}x_1 \\ a_{22}x_2 \end{bmatrix} = a_{11}x_1^2 + a_{22}x_2^2$$

与 $x'x$ 的简单平方和不同，此式给出了一个"加权"平方和。完全相同的结果源于

$$(x'A)x = [a_{11}x_1 \quad a_{22}x_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = a_{11}x_1^2 + a_{22}x_2^2$$

矩阵乘法也服从分配律。

**分配律**　　$A(B+C)=AB+AC$　　以 $A$ 左乘

$(B+C)A=BA+CA$　　以 $A$ 右乘

在每一情况下,可加性条件和可相乘条件均得满足。

## 练习 4.4

1　已知 $A=\begin{bmatrix} 3 & 6 \\ 2 & 4 \end{bmatrix}$,$B=\begin{bmatrix} -1 & 7 \\ 8 & 4 \end{bmatrix}$,和 $C=\begin{bmatrix} 3 & 4 \\ 1 & 9 \end{bmatrix}$,证明

(a) $(A+B)+C=A+(B+C)$

(b) $(A+B)-C=A+(B-C)$

2　减去矩阵 $B$ 可视作加上矩阵 $(-1)B$。加法的交换律允许我们作下述表述吗?

$A-B=B-A$

若不能,如何作出正确表述?

3　用下列矩阵验证乘法的结合律:

$$A=\begin{bmatrix} 5 & 3 \\ 0 & 5 \end{bmatrix} \quad B=\begin{bmatrix} -8 & 0 & 7 \\ 1 & 3 & 2 \end{bmatrix} \quad C=\begin{bmatrix} 1 & 0 \\ 0 & 3 \\ 7 & 1 \end{bmatrix}$$

79

4　对任意两个标量 $g$ 和 $k$,证明下式成立:

(a) $k(A+B)=kA+kB$

(b) $(g+k)A=gA+kA$

5　证明:$(A+B)(C+D)=AC+AD+BC+BD$。

6　若例 5 中的矩阵 $A$ 的全部四个元素为非零,$x'Ax$ 仍能给出加权平方和吗?结合律仍适用吗?

## 4.5　单位矩阵与零矩阵

### 单位矩阵

前面已经提到**单位矩阵**这一术语。它被定义为主对角线元素为 1,其余元素为零的**方阵**。单位矩阵以符号 $I$ 或 $I_n$ 表示,下标 $n$ 标示行或列数。因此,

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

但二者也可以 $I$ 表示。

这种特殊类型的单位矩阵之所以重要,是因为它与标量代数中的 1 起着类似的作用。对于任意数 $a$,我们有 $1(a)=a(1)=a$。类似地,对于任意矩阵 $A$,我们有

$$IA = AI = A \tag{4.8}$$

**例 1**　令 $A=\begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \end{bmatrix}$,则

$$IA = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \end{bmatrix} = A$$

$$AI = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \end{bmatrix} = A$$

因为 $A$ 为 $2\times3$ 矩阵,$A$ 被 $I$ 左乘和右乘分别要求不同维数的单位

矩阵,即 $I_2$ 和 $I_3$。但若 $A$ 为 $n \times n$ 矩阵,则可用相同单位矩阵 $I_n$,使得(4.8)变为 $I_nA = AI_n$,这是矩阵乘法不服从交换律的一个例外。

在乘积过程中,单位矩阵的特殊性质使得可以**插入或删除**一个单位矩阵而不影响矩阵积。由(4.8)可直接得出此结论。回忆结合律,我们有,比如

$$\underset{(m\times n)}{A}\ \underset{(n\times n)}{I}\ \underset{(n\times p)}{B} = (AI)B = \underset{(m\times n)}{A}\ \underset{(n\times p)}{B}$$

80 此式表明单位矩阵 $I$ 的存在与否并不影响矩阵积。注意,无论 $I$ 是否出现在矩阵积中,相乘的维数相容性条件仍要保留。

当 $A = I_n$ 时,(4.8)出现一种令人感兴趣的情况。因为这时,我们有:

$$AI_n = (I_n)^2 = I_n$$

此式表明单位矩阵的平方仍等于其自身。此结果可一般化为

$$(I_n)^k = I_n(k = 1,2,\cdots)$$

即单位矩阵无论自乘多少次仍保持不变。具有这种性质(即 $AA = A$)的矩阵称作**幂等矩阵**。

### 零矩阵

恰如单位矩阵 $I$ 与数 1 起同样作用一样,以“0”表示的零矩阵,与数 0 起着同样的作用。零矩阵是一个所有元素均为零的简单矩阵。不同于 $I$,零矩阵并不仅限于方阵。因此它可能写成:

$$\underset{(2\times 2)}{0} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \text{ 和 } \underset{(2\times 3)}{0} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

等等。零方阵为幂等矩阵,但不为方阵的零矩阵则不是幂等的(为什么?)。

作为与数0作用相似的矩阵,零矩阵的加法和乘法服从下列运算法则(在服从可相加、相乘的前提下):

$$\underset{(m\times n)}{A} + \underset{(m\times n)}{0} = \underset{(m\times n)}{0} + \underset{(m\times n)}{A} = \underset{(m\times n)}{A}$$

$$\underset{(m\times n)(n\times p)}{A \quad 0} = \underset{(m\times p)}{0} \text{ 和 } \underset{(q\times m)(m\times n)}{0 \quad A} = \underset{(q\times n)}{0}$$

注意,在乘法中,等号左边矩阵和等号右边矩阵的维数可能不同。

**例2**

$$A + 0 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = A$$

**例3**

$$\underset{(2\times3)}{A} \; \underset{(3\times1)}{0} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \underset{(2\times1)}{0}$$

左边的零矩阵为3×1零向量,右边为2×1零向量。

## 矩阵代数的特征

81

尽管矩阵代数与标量代数有明显的相似之处,但矩阵代数确实表现出一些独特的特点,使我们不能毫无疑问地"照搬"标量代数的运算规律。我们已经知道,在矩阵代数中,一般 $AB \neq BA$。现在我们考察矩阵代数的另外两个特征。

首先,在标量情况下,$ab=0$ 总是意味着或者 $a$ 为零,或者 $b$ 为零,但在矩阵乘法中并不如此。因为我们有:

$$AB = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -2 & 4 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0$$

无论 $A$ 还是 $B$ 都不是零矩阵。

另外，在标量(数)情况下，方程 $cd = ce(c \neq 0)$ 意味着 $d = e$。但对矩阵而言并非如此。因而，已知

$$C = \begin{bmatrix} 2 & 3 \\ 6 & 9 \end{bmatrix} \qquad D = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \qquad E = \begin{bmatrix} -2 & 1 \\ 3 & 2 \end{bmatrix}$$

我们求出

$$CD = CE = \begin{bmatrix} 5 & 8 \\ 15 & 24 \end{bmatrix}$$

但是 $D \neq E$。

这些奇特结果实际上仅与被称作奇异矩阵的一类特殊矩阵有关，矩阵 $A$、$B$、$C$ 就是这类矩阵的一些例子。(粗略地讲，这些矩阵有一行是另一行的倍数。)无论如何，这些例子确实揭示了将代数定理不加思考地运用到矩阵运算的一些弊端。

## 练习 4.5

---

给定 $A = \begin{bmatrix} -1 & 8 & 7 \\ 0 & -2 & 4 \end{bmatrix}$，$b = \begin{bmatrix} 9 \\ 6 \\ 0 \end{bmatrix}$，$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$：

1 计算：(*a*) $AI$ (*b*) $IA$ (*c*) $Ix$ (*d*) $x'I$

指明每一情况下所应用单位矩阵的维数。

2 计算：(*a*) $Ab$ (*b*) $AIb$ (*c*) $x'IA$ (*d*) $x'A$

在(*b*)中插入单位矩阵 $I$ 影响(*a*)的结果吗？在(*d*)中删除 $I$ 影响在(*c*)中的结果吗？

3 由下列所得的各零矩阵的维数为多少？

(*a*) $4 \times 2$ 零矩阵左乘 $A$；

(*b*) $3 \times 6$ 零矩阵右乘 $A$；

(c) 4×3 零矩阵左乘 $b$；

(d) 1×5 零矩阵右乘 $x$。



4 证明对角阵，即下述形式的矩阵

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

仅当对角线上每一元素为 1 或 0 时，才是幂等矩阵。由上述矩阵可构造多少个不同数字的幂等矩阵？

---

## 4.6 矩阵的转置与逆

当矩阵 $A$ 的行和列互换——使得第一行变为第一列，反之亦然——我们得到 $A$ 的**转置**，以 $A'$ 或 $A^T$ 表示。符号"′"我们并不陌生，在区分行向量和列向量时，我们曾使用过它。用这里新引进的术语，行向量 $x'$ 构成列向量 $x$ 的转置。显然，另外一种表示转置的符号中，$T$ 是 *Transpose*（转置）的简写。

**例 1** 给定 $\underset{(2\times 3)}{A} = \begin{bmatrix} 3 & 8 & -9 \\ 1 & 0 & 4 \end{bmatrix}$ 和 $\underset{(2\times 2)}{B} = \begin{bmatrix} 3 & 4 \\ 1 & 7 \end{bmatrix}$

我们将行和列互换并写成

$$\underset{(3\times 2)}{A'} = \begin{bmatrix} 3 & 1 \\ 8 & 0 \\ -9 & 4 \end{bmatrix} \text{和} \underset{(2\times 2)}{B'} = \begin{bmatrix} 3 & 1 \\ 4 & 7 \end{bmatrix}$$

根据定义，若 $A$ 为 $m \times n$ 矩阵，则其转置必定为 $n \times m$ 矩阵。

**例 2** 如果 $C=\begin{bmatrix}9 & -1\\2 & 0\end{bmatrix}$ 和 $D=\begin{bmatrix}1 & 0 & 4\\0 & 3 & 7\\4 & 7 & 2\end{bmatrix}$,那么

$$C'=\begin{bmatrix}9 & 2\\-1 & 0\end{bmatrix} \text{和} D'=\begin{bmatrix}1 & 0 & 4\\0 & 3 & 7\\4 & 7 & 2\end{bmatrix}$$

这里,每一转置矩阵的维数与原矩阵是相同的。

在 $D'$中,我们还要注意其极为显著的一个结果:它不仅保留了原矩阵的维数,而且也保留了原矩阵的排列。$D=D'$是由于元素对主对角线对称的结果。将 $D$ 中的主对角线视作镜子,位于其东北方向的元素恰好为其西南方向元素的像;因而第一行的元素与第一列相同,等等。矩阵 $D$ 就是这类被称作对称矩阵的方阵的一个例子。另一个对称矩阵的例子是单位矩阵 $I$,它像对称矩阵一样,具有转置 $I'=I$。

83

## 转置的性质

转置具有下列性质:

$$(A')' = A \tag{4.9}$$

$$(A+B)' = A'+B' \tag{4.10}$$

$$(AB)' = B'A' \tag{4.11}$$

第一个性质是说,转置矩阵的转置为原矩阵——一个不言自明的结论。

第二个性质可用文字这样表述:和的转置为转置的和。

**例 3** 若 $A=\begin{bmatrix}4 & 1\\9 & 0\end{bmatrix}$, $B=\begin{bmatrix}2 & 0\\7 & 1\end{bmatrix}$,则

$$(A+B)' = \begin{bmatrix} 6 & 1 \\ 16 & 1 \end{bmatrix}' = \begin{bmatrix} 6 & 16 \\ 1 & 1 \end{bmatrix}$$

和 $$A' + B' = \begin{bmatrix} 4 & 9 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 2 & 7 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 16 \\ 1 & 1 \end{bmatrix}$$

第三个性质为乘积的转置是转置的乘积，但转置矩阵顺序颠倒。为理解顺序颠倒的重要性，我们考察(4.11)左右两边两个乘积维数的可相乘条件。若我们令 $A$ 为 $m \times n$ 矩阵，$B$ 为 $n \times p$ 矩阵，则 $AB$ 为 $m \times p$ 矩阵，$(AB)'$ 将是 $p \times m$ 矩阵。要使等式成立，右边表达式 $B'A'$ 必有相等的维数。因 $B'$ 为 $p \times n$，$A'$ 为 $n \times m$，其积 $B'A'$ 确实为 $p \times m$，正如所要求的那样。这样 $B'A'$ 的维数就得到预期的结果。另外还要注意，除非 $m = p$，否则 $A'B'$ 甚至无定义。

**例 4** 已知 $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ 和 $B = \begin{bmatrix} 0 & -1 \\ 6 & 7 \end{bmatrix}$，具有

$$(AB)' = \begin{bmatrix} 12 & 13 \\ 24 & 25 \end{bmatrix}' = \begin{bmatrix} 12 & 24 \\ 13 & 25 \end{bmatrix}$$

和 $$B'A' = \begin{bmatrix} 0 & 6 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 12 & 24 \\ 13 & 25 \end{bmatrix}$$

证明了该性质。

### 逆及其性质

给定矩阵 $A$，总可以推导出其转置 $A'$，另一方面，其逆矩阵——另一类导出的矩阵——可能存在，也可能不存在。矩阵 $A$ 84 的逆，以 $A^{-1}$ 表示，仅当 $A$ 为方阵时才有定义。在此情况下逆为满足下列条件的矩阵：

$$AA^{-1} = A^{-1}A = I \tag{4.12}$$

即无论 $A$ 是被 $A^{-1}$左乘还是右乘,其积总是同一单位矩阵。这是矩阵乘法不服从交换律的另一例外。

以下各点值得注意:

1. 并非每一方阵均有逆阵,方阵是逆阵存在的必要条件,但并非充分条件。若方阵 $A$ 有逆阵,则 $A$ 被称作非奇异的;若 $A$ 无逆阵,则 $A$ 为奇异矩阵。

2. 在 $A^{-1}$存在,则 $A$ 可视为 $A^{-1}$的逆,恰如 $A^{-1}$可视为 $A$ 的逆一样。简言之,$A$ 与 $A^{-1}$互为逆矩阵。

3. 若 $A$ 为 $n \times n$ 矩阵,则 $A^{-1}$必为 $n \times n$ 矩阵,否则,它不能既满足左乘又满足右乘的可相乘条件。乘积而得到的单位矩阵也是 $n \times n$ 矩阵。

4. 若逆阵存在,则它是唯一的。为证明逆的惟一性,我们假定已知 $B$ 为 $A$ 的逆,所以

   $$AB = BA = I$$

   现在假设存在另一矩阵 $C$ 使得 $AC = CA = I$,通过将 $AB = I$ 两边同时以 $C$ 左乘,我们得到

   $$CAB = CI(= C) \qquad [由(4.8)]$$

   因为由假设 $CA = I$,上一方程可简化为

   $$IB = C \text{ 或 } B = C$$

   即 $B$ 与 $C$ 必然相同,是同一逆矩阵。因此我们说 $A$ 的逆,而不说 $A$ 的一个逆。

5. 条件(4.12)的两部分,即 $AA^{-1} = I$ 和 $A^{-1}A = I$ 实际上互相包容,所以满足一个方程就足以确定 $A$ 与 $A^{-1}$的逆的关系。为证明这一点,我们需证明若 $AA^{-1} = I$,若存在

矩阵$B$使得$BA=I$，则$B=A^{-1}$（所以$BA=I$实际上就是$A^{-1}A=I$）。我们以$A^{-1}$同时右乘给定方程$BA=I$，则

$$(BA)A^{-1}=IA^{-1}$$

$$B(AA^{-1})=IA^{-1}\text{［结合律］}$$

$$BI=IA^{-1}\text{［由假设 }AA^{-1}=I\text{］}$$

因而，正如要求的那样，

$$B=A^{-1}\qquad\text{［由(4.8)］}$$

85

类似地，还可以证明，若$A^{-1}A=I$，则产生$CA^{-1}=I$的惟一矩阵$C$是$C=A$。

**例5** 令$A=\begin{bmatrix}3&1\\0&2\end{bmatrix}$和$B=\frac{1}{6}\begin{bmatrix}2&-1\\0&3\end{bmatrix}$，则因$B$中的标量乘子$(\frac{1}{6})$可移至后面（交换律），我们可以写成：

$$AB=\begin{bmatrix}3&1\\0&2\end{bmatrix}\begin{bmatrix}2&-1\\0&3\end{bmatrix}\frac{1}{6}=\begin{bmatrix}6&0\\0&6\end{bmatrix}\frac{1}{6}=\begin{bmatrix}1&0\\0&1\end{bmatrix}$$

这就确立了$B$为$A$的逆，反之亦然。颠倒乘积顺序，正如我们预期的那样，也得到同样的单位矩阵：

$$BA=\frac{1}{6}\begin{bmatrix}2&-1\\0&3\end{bmatrix}\begin{bmatrix}3&1\\0&2\end{bmatrix}=\frac{1}{6}\begin{bmatrix}6&0\\0&6\end{bmatrix}=\begin{bmatrix}1&0\\0&1\end{bmatrix}$$

逆阵的下述三个性质是令人感兴趣的。若$A$和$B$为两个$n\times n$非奇异矩阵，则

$$(A^{-1})^{-1}=A\qquad(4.13)$$

$$(AB)^{-1}=B^{-1}A^{-1}\qquad(4.14)$$

$$(A')^{-1}=(A^{-1})'\qquad(4.15)$$

第一个性质表明逆阵的逆就是原矩阵。第二个性质表明积的逆矩阵等于逆矩阵的积，但乘积顺序颠倒。最后一个性质意味着转置

的逆为逆的转置。应注意，在这些表述中，逆的存在和可相乘条件是预先假设的。

(4.13)的有效性是显而易见的，所以我们仅证明(4.14)和(4.15)。给定积 $AB$，我们求出其逆，称之为 $C$。由(4.12)我们知道 $CAB=I$；因此，以 $B^{-1}A^{-1}$右乘两边，将产生

$$CABB^{-1}A^{-1} = IB^{-1}A^{-1}(= B^{-1}A^{-1}) \tag{4.16}$$

但左边可简化为

$$\begin{aligned} CA(BB^{-1})A^{-1} &= CAIA^{-1} && [\text{由}(4.12)] \\ &= CAA^{-1} = CI = C && [\text{由}(4.12)\text{和}(4.8)] \end{aligned}$$

将其代入(4.16)则有 $C=B^{-1}A^{-1}$，或者换句话说，$AB$ 的逆等于 $B^{-1}A^{-1}$，这正是前面所宣称的结果。在这个证明中，方程 $AA^{-1}=A^{-1}A=I$ 被运用了两次。注意，当且仅当一个矩阵和它的逆在积中严格毗邻时，才可以应用这个方程。我们可以写 $AA^{-1}B=IB=B$，但绝不能写 $ABA^{-1}=B$。

(4.15)的证明如下。给定 $A'$，我们求其逆，称作 $D$。由定义，则我们有 $DA'=I$。但我们知道：

$$(AA^{-1})' = I' = I$$

86 产生同样的单位矩阵。因此我们可以写出

$$\begin{aligned} DA' &= (AA^{-1})' \\ &= (A^{-1})'A' && [\text{由}(4.11)] \end{aligned}$$

两边以$(A')^{-1}$右乘，得到

$$DA'(A')^{-1} = (A^{-1})'A'(A')^{-1}$$

或 $$D = (A^{-1})' \qquad [\text{由}(4.12)]$$

因此 $A'$的逆等于$(A^{-1})'$，正如所声明的那样。

在刚才的证明中，数学运算是对整个数据进行的，如果不把这

些数块作为数学整体(矩阵),则同样的运算会变得极其冗长复杂。矩阵代数的精妙之处就在于运算的简化。

## 逆矩阵与线性方程组的解

逆矩阵的概念可直接方便地应用于解联立方程组。关于(4.3)中的方程组,我们早已指出,它用矩阵的概念可以写成

$$\underset{(3\times3)}{A}\ \underset{(3\times1)}{x} = \underset{(3\times1)}{d} \tag{4.17}$$

其中 $A$,$x$ 和 $d$ 在(4.4)中已有定义,现在,若 $A^{-1}$存在,方程(4.17)两边均以 $A^{-1}$左乘,将产生

$$A^{-1}Ax = A^{-1}d$$

或 (4.18)

$$\underset{(3\times1)}{x} = \underset{(3\times3)}{A^{-1}}\ \underset{(3\times1)}{d}$$

(4.18)的左侧为变量的列向量,而右侧积则是某些已知数的列向量。因此,由矩阵或向量相等的定义,(4.18)表明了满足方程组变量的一组值,即解值。而且,若 $A^{-1}$存在,则它是唯一的,所以 $A^{-1}d$一定是解值的唯一向量。因而我们可以将(4.18)中的 $x$ 向量写成 $\bar{x}$,以表示其作为唯一解的地位。

检验逆存在及计算逆的方法将在下一章中讨论。但这里可以表明(4.4)中矩阵 $A$ 的逆为

$$A^{-1} = \frac{1}{52}\begin{bmatrix} 18 & -16 & -10 \\ -13 & 26 & 13 \\ -17 & 18 & 21 \end{bmatrix}$$

因而(4.18)的结果是

$$\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{bmatrix} = \frac{1}{52}\begin{bmatrix} 18 & -16 & -10 \\ -13 & 26 & 13 \\ -17 & 18 & 21 \end{bmatrix}\begin{bmatrix} 22 \\ 12 \\ 10 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

87

此式给出解值 $\bar{x}_1 = 2$，$\bar{x}_2 = 3$，$\bar{x}_3 = 1$。

上式作为求解线性方程组 $Ax = d$(其中系数矩阵 $A$ 为非奇异阵)的一种方法,其要点是:首先求出逆阵 $A^{-1}$,然后以常向量 $d$ 右乘 $A^{-1}$。则积 $A^{-1}d$ 将给出变量的解值。

## 练习 4.6

---

1 给定 $A = \begin{bmatrix} 2 & 4 \\ -1 & 3 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 8 \\ 0 & 1 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 0 & 9 \\ 6 & 1 & 1 \end{bmatrix}$,求 $A'$,$B'$,$C'$。

2 运用上题中给出的矩阵验证

$(a)(A+B)' = A' + B'$ $(b)(AC)' = C'A'$

3 证明任意可相乘矩阵 $A$、$B$、$C$,方程$(ABC)' = C'B'A'$成立,从而将(4.11)的结论推广到三个矩阵相乘的情况。

4 给定下列四个矩阵,检验是否其中任意一个矩阵均是其它矩阵的逆。

$$D = \begin{bmatrix} 1 & 12 \\ 0 & 3 \end{bmatrix} \quad E = \begin{bmatrix} 1 & 1 \\ 6 & 8 \end{bmatrix} \quad F = \begin{bmatrix} 1 & -4 \\ 0 & \frac{1}{3} \end{bmatrix} \quad G = \begin{bmatrix} 4 & -\frac{1}{2} \\ -3 & -\frac{1}{2} \end{bmatrix}$$

5 证明对任意可相乘非奇异矩阵 $A,B,C$,方程$(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$,从而将(4.14)的结论推广。

6 令 $A = I - X(X'X)^{-1}X'$.

(*a*) $A$ 必须为方阵吗?$(X'X)$必须为方阵吗?$X$ 必须为方阵吗?

(*b*) 证明矩阵 $A$ 为幂等矩阵。[提示:若 $X'$与 $X$ 不为方阵,则不适于应用(4.14)。]

---

# 第5章 线性模型与矩阵代数(续)

88

第4章已经表明,无论多大的线性方程组,均可用简洁的矩阵符号来表示。而且若方程组系数矩阵的逆存在,还可通过求系数矩阵逆矩阵的方法解方程组。我们现在则必须考虑解决这些问题:如何检验逆矩阵的存在,以及如何求逆矩阵。只有我们回答了这些问题,我们才能将矩阵代数应用于经济模型。

## 5.1 矩阵非奇异性的条件

仅当系数矩阵 $A$ 为方阵时,它才可能有逆矩阵(即可能是“非奇异的”)。但正如前面已指出的那样,方阵条件是逆矩阵 $A^{-1}$ 存在的必要条件,而非充分条件。一个矩阵可能是方阵,但也可能是奇异矩阵(没有逆矩阵)。

### 必要条件和充分条件

在经济学中,我们经常用到“必要条件”和“充分条件”这些概念。在进一步讨论前,理解其确切含义是重要的。

必要条件具有先决条件的性质:假设仅当另一陈述 $q$ 为真时,陈述 $p$ 才为真,则 $q$ 便构成 $p$ 的必要条件。我们可用符号表示如下: 89

$$p \Rightarrow q \tag{5.1}$$

读作"仅当 $q$ 成立时，$p$ 成立"，或者"若 $p$ 成立，则 $q$ 成立"。将(5.1)解释成"$p$ 成立意味着 $q$ 成立"在逻辑上也是正确的。当然，也可能出现这种情况，若我们同时还有 $p \Rightarrow w$，则 $q$ 和 $w$ 都是 $p$ 的必要条件。

**例 1** 我们令 $p$ 为陈述"某人是一个父亲"，$q$ 为陈述"某人是男性"，则逻辑表述 $p \Rightarrow q$ 成立，因为仅当一个人为男性时他才能是一个父亲，男性是作父亲的必要条件。但要注意，反之不成立；父亲身份不是男性的必要条件。

另一种不同类型的情况是若 $q$ 为真，则 $p$ 为真，但 $q$ 不为真时 $p$ 也可能为真。在此情况下，$q$ 被称作 $p$ 的充分条件。$q$ 为真足以保证 $p$ 为真成立，但它不是 $p$ 为真的必要条件。以符号表示这种情况：

$$p \Leftarrow q \tag{5.2}$$

读作："当 $q$ 成立，则 $p$ 成立"(没有"仅"这个字)，或"若 $q$ 成立，则 $p$ 成立"，好像反读(5.2)式。还可以将其解释为"$q$ 成立意味着 $p$ 成立"。

**例 2** 若我们令 $p$ 代表陈述"一个人可到达欧洲"，$q$ 代表陈述"一个人乘飞机到达欧洲"，则 $p \Leftarrow q$。乘飞机可抵达欧洲，但海上交通也可抵达欧洲，所以乘飞机并非先决条件。因而我们可以写 $p \Leftarrow q$，但不能写 $p \Rightarrow q$。

在第三种可能的情况中，$q$ 既是 $p$ 的必要条件，也是充分条件。在此情况下，我们写成：

$$p \Leftrightarrow q \tag{5.3}$$

读作："当且仅当 $q$ 成立，则 $p$ 成立"，双箭头实际上是(5.1)和

(5.2)中箭头的结合，因而同时使用“当”和“仅当”两个词。注意，(5.3)表明，不仅 $p$ 成立意味着 $q$ 成立，而且 $q$ 成立意味着 $p$ 成立。

**例3** 若我们令 $p$ 代表陈述“少于30天的一个月”，而 $q$ 代表陈述“它是二月”，则 $p \Leftrightarrow q$。少于30天的月份，必然是2月。反过来，2月份的设定足以确定该月少于30天。因而 $q$ 为 $p$ 的充分必要条件。

为证明 $p \Rightarrow q$，需证明在逻辑上 $q$ 可从 $p$ 推出。类似地，为证明 $p \Leftarrow q$，需证明在逻辑上 $p$ 可从 $q$ 推出，但要证明 $p \Leftrightarrow q$，则需证明 $p$ 和 $q$ 可互相从对方推出。

90

## 非奇异性条件

当方阵条件已满足时，矩阵非奇异性的充分条件是行线性无关(或者同样的，列线性无关)。方阵条件和线性无关条件结合在一起，构成了非奇异性的充分必要条件(非奇异性⇔方阵和线性无关)。

一个 $n \times n$ 系数矩阵 $A$ 可以视为行向量的有序集，即视为一个元素为行向量的列向量：

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} v'_1 \\ v'_2 \\ \vdots \\ v'_n \end{bmatrix}$$

其中 $v_i' = [a_{i1} \ a_{i2} \cdots \ a_{in}]$，$i = 1, 2, \cdots, n$。因为行向量线性无关，所以没有一行为其它行的线性组合。更正式地，如在4.3节所提到的那样，行线性无关要求标量 $k_i$ 的集合满足下列向量方程

$$\sum_{i=1}^{n} k_i v'_i = \underset{(1\times n)}{0} \tag{5.4}$$

对所有的 $i, k_i = 0$。

**例 4** 若系数矩阵

$$A = \begin{bmatrix} 3 & 4 & 5 \\ 0 & 1 & 2 \\ 6 & 8 & 10 \end{bmatrix} = \begin{bmatrix} v'_1 \\ v'_2 \\ v'_3 \end{bmatrix}$$

则因[6 8 10] = 2 [3 4 5]，我们有 $v_3' = 2v_1' = 2v_1' + 0v_2'$。所以第三行可以表示为前两行的线性组合，行并非线性无关。因此，我们可以将方程另写成

$$2v'_1 + 0v'_2 - v'_3 = [6\ 8\ 10] + [0\ 0\ 0] - [6\ 8\ 10] = [0\ 0\ 0]$$

由于导向(5.4)零向量的标量集 $k_i$ 对所有的 $i$ 并不都全为零，所以行向量是线性相关的。

91 不同于方阵条件，线性无关条件一般难以看一眼就确定。所以需开发检验行或列线性无关的方法。但在我们着手这个任务之前，我们首先需要对为什么线性无关条件和方型条件联系在一起有一个直观的了解。由 3.4 节对计算方程数与未知数的讨论，我们可以回忆起这个一般结论：要使方程组有唯一解，仅满足方程数与未知数个数相同的条件是不够的，此外还要有方程必须彼此相容且函数无关（在现在线性方程组的情况下，意味着线性无关）。显然，在“方程数与未知数个数相同”这一标准与系数矩阵为方阵（行数与列数相同）这一情况之间存在着紧密的联系。“行间线性无关”的要求确实可以排除方程间的不相容和线性相关。因此，将其结合在一起，系数矩阵的方形和行线性无关这两个要求相当于 3.4 节所述的方程存在惟一解的条件。

我们举例说明系数矩阵行间线性相关如何导致方程之间的不相容或线性相关。令方程组 $Ax=d$ 取下述形式

$$\begin{bmatrix}10 & 4\\5 & 2\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}d_1\\d_2\end{bmatrix}$$

其中系数矩阵 $A$ 包含线性相关的行：$v_1'=2v_2'$。（注意，矩阵的列也是线性相关的，第一列为第二列的 5/2。）我们未设定常数项 $d_1$ 和 $d_2$ 的值，但关于其相对值仅有两种明显的可能性：(1) $d_1=2d_2$ 和 $d_1\neq 2d_2$。在第一种情况下，比如 $d_1=12$，$d_2=6$，两个方程是相容的但线性相关（恰如矩阵 $A$ 的两行线性相关一样），因为第一个方程等于第二个方程乘 2。这时一个方程足矣，方程组实际上简化为一个方程 $5x_1+2x_2=6$，它具有无数个解。在第二种可能的情况下，比如 $d_1=12$，$d_2=0$，两个方程是不相容的，因为若第一个方程（$10x_1+4x_2=12$）成立，对每一项以 2 相除，我们将其化简为：$5x_1+2x_2=6$；结果第二个方程（$5x_1+2x_2=0$）不可能同时成立，因此方程无解。

结果是只要系数矩阵的行线性相关，就没有适当的唯一解存在（在上述任意一种情况下），事实上，获得唯一解的唯一方式是系数矩阵的行或列线性无关。在此情况下，$A$ 为非奇异矩阵，这意味着逆 $A^{-1}$ 确实存在。且唯一解 $\bar{x}=A^{-1}d$ 可求。

## 矩阵的秩

尽管行线性无关的概念仅在与方阵有关的情形下讨论，但它 92 同样可以运用于任意 $m\times n$ 矩阵。在一个矩阵中，若线性无关的最大行数为 $\gamma$，则称该矩阵的秩为 $r$。（秩也告诉我们该矩阵线性

无关的最大列数。)$m \times n$ 矩阵的秩的最大值等于 $m$、$n$ 中较小的一个。

由定义,一个 $n \times n$ 非奇异矩阵 $A$ 具有 $n$ 个线性无关的行或列,所以其秩必为 $n$。反之,一个秩为 $n$ 的 $n \times n$ 矩阵必为非奇异矩阵。

## 练习 5.1

---

1 在下列每对陈述中,令 $p$ 代表第一个陈述,$q$ 代表第二个陈述。指明下列每一情况是适于(5.1),(5.2)还是(5.3)。

($a$) 今天是假期;今天是感恩节。

($b$) 一个几何图形有四条边;它是一个矩形。

($c$) 两个有序偶$(a,b)$,$(b,a)$相等;$a$ 等于 $b$。

($d$) 一个数为有理数;一个数可以表示为两个整数的比。

($e$) 一个 4×4 矩阵为非奇异矩阵;该矩阵的秩为 4。

($f$) 我的汽车油箱是空的;我不能起动我的汽车。

($g$) 因邮资不足,这封信被返还寄信人;寄信人忘记在信封上贴邮票。

2 令 $p$ 表示陈述"一个几何图形为正方形";$q$ 代表如下陈述:

($a$) 该几何图形有四条边。

($b$) 该几何图形有四条等边。

($c$) 几何图形有四条相等的边,且每条边与邻边垂直。

对于每一陈述,$p \Rightarrow q$, $p \Leftarrow q$, $p \Leftrightarrow q$ 哪种情况为真?

3 在下面每一矩阵中,行向量均线性无关吗?

($a$) $\begin{bmatrix} 1 & 8 \\ 9 & -3 \end{bmatrix}$ ($b$) $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ ($c$) $\begin{bmatrix} 0 & 2 \\ 3 & 2 \end{bmatrix}$ ($d$) $\begin{bmatrix} -1 & 4 \\ 2 & -8 \end{bmatrix}$

4 检验上题的列是否也都线性无关。你得出的结论与行线性无关相同吗?

---

## 5.2 用行列式检验非奇异性

要确定一个方阵是否为非奇异矩阵，我们可以运用行列式的概念。

### 行列式与非奇异性

93

方阵 $A$ 的行列式以 $|A|$ 表示，它是唯一定义的与该矩阵相联系的标量(数)。只有对方阵才能定义行列式。对于一个 $2\times 2$ 矩阵 $A=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$，其行列式定义为下面两项的和：

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12} [\text{一个标量}] \tag{5.5}$$

将 $A$ 的主对角线上的两个元素相乘，然后减去其余两元素之积，便得到上式。根据矩阵 $A$ 的维数，(5.5)所定义的 $|A|$ 被称作二阶行列式。

**例1** 已知 $A=\begin{bmatrix} 10 & 4 \\ 8 & 5 \end{bmatrix}$，$B=\begin{bmatrix} 3 & 5 \\ 0 & -1 \end{bmatrix}$ 其行列式为

$$|A| = \begin{vmatrix} 10 & 4 \\ 8 & 5 \end{vmatrix} = 10(5) - 8(4) = 18$$

和 $$|B| = \begin{vmatrix} 3 & 5 \\ 0 & -1 \end{vmatrix} = 3(-1) - 0(5) = -3$$

尽管由定义，一个行列式(以两条纵线括住，而不用方括号括住)是一个标量，但这样的矩阵本身并没有一个数值。换句话说，一个行

列式可化简为一个数，而矩阵则相反，它是一个数块的整体。还需强调的是，行列式仅对方阵有定义，而矩阵则不要求一定为方阵。

即使在讨论的初期，我们也可能对矩阵 $A$ 行线性相关与行列式 $|A|$ 之间的关系，形成模糊的印象。两个矩阵

$$C = \begin{bmatrix} c'_1 \\ c'_2 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 3 & 8 \end{bmatrix} \quad 和 \quad D = \begin{bmatrix} d'_1 \\ d'_2 \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 8 & 24 \end{bmatrix}$$

均为行线性相关，因为 $c_1' = c_2'$，$d_2' = 4d_1'$。两个矩阵的行列式的结果都等于 0：

$$|C| = \begin{vmatrix} 3 & 8 \\ 3 & 8 \end{vmatrix} = 3(8) - 3(8) = 0$$

$$|D| = \begin{vmatrix} 2 & 6 \\ 8 & 24 \end{vmatrix} = 2(24) - 8(6) = 0$$

此结果明显表明“成零”行列式（零值行列式）与线性相关具有某种联系。我们将看到事实确实如此。因而，行列式 $|A|$ 的值不仅可以作为检验矩阵 $A$ 行线性无关（因而为非奇异矩阵）的一个标准，而且还可以作为逆 $A^{-1}$ 的计算结果，若逆存在的话。

94

但首先我们必须讨论高阶行列式以扩大我们的视野。

## 计算三阶行列式

三阶行列式与一个 $3\times3$ 矩阵相联系。给定

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

其行列式的值为

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

$$= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \quad [= \text{一个标量}] \qquad (5.6)$$

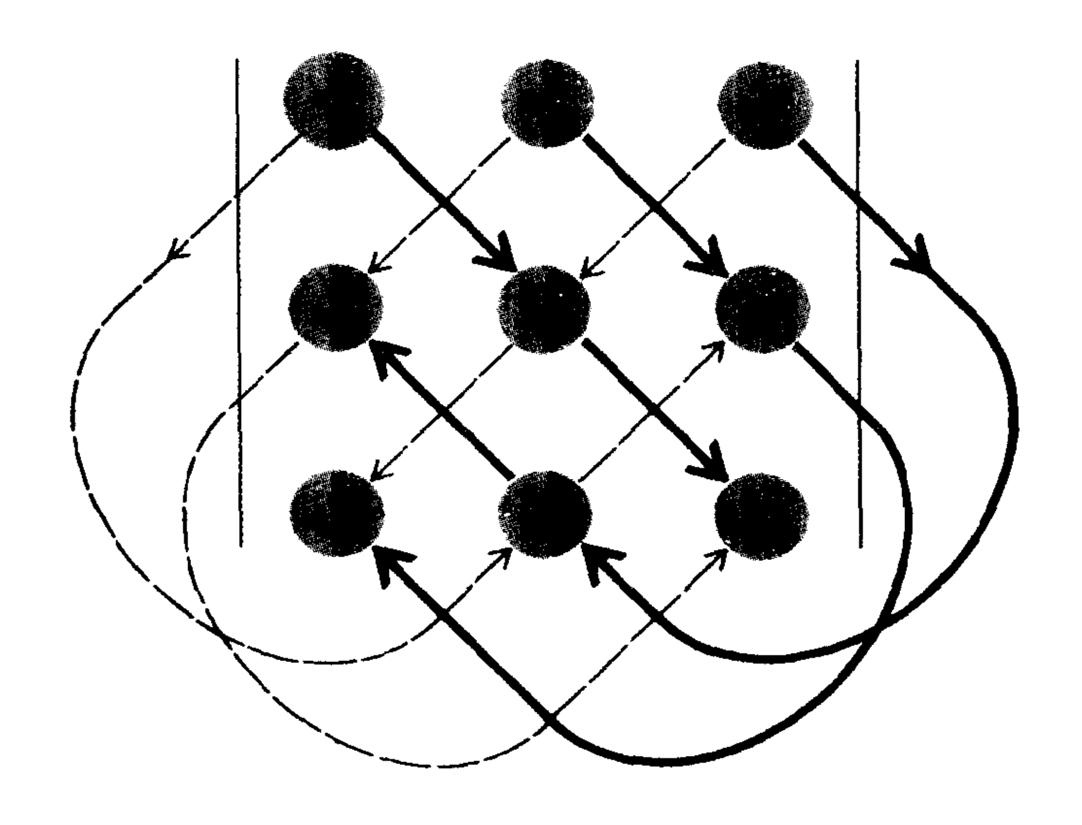


图 5.1

首先，观察(5.6)式的下面两行，我们看到 $|A|$ 的值表示为六项乘积之和，其中三项前面为负号，三项前面为正号。尽管此和看起来有些复杂，仍然有一个简单的方法从一个给定的三阶行列式中“捕捉到”这六项。这在图 5.1 中得到了最好的解释。在图 5.1 所示的行列式中，最上一行的每一元素用两个实箭头以下述 95

方式与另外两个元素相联：$a_{11} \to a_{22} \to a_{33}$，$a_{12} \to a_{23} \to a_{31}$；$a_{13} \to a_{32} \to a_{21}$。每三个这样的元素通过相乘联系起来，其积便是(5.6)六项乘积中的一项。实箭头积项前面的符号为正。

另外，最上一行每一元素还用两个虚箭头与另外两个元素相联：$a_{11} \to a_{32} \to a_{23}$，$a_{12} \to a_{21} \to a_{33}$，$a_{13} \to a_{22} \to a_{31}$。每三个元素也以乘积结合在一起，此积也是(5.6)中六项之一项。这些积前面的符号为负。全部六项乘积之和便是行列式的和。

**例 2**

$$\begin{vmatrix} 2 & 1 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = (2)(5)(9) + (1)(6)(7) + (3)(8)(4)$$

$$-(2)(8)(6) - (1)(4)(9) - (3)(5)(7) = -9$$

**例 3**

$$\begin{vmatrix} -7 & 0 & 3 \\ 9 & 1 & 4 \\ 0 & 6 & 5 \end{vmatrix} = (-7)(1)(5) + (0)(4)(0) + (3)(6)(9)$$

$$-(-7)(6)(4) - (0)(9)(5) - (3)(1)(0) = 295$$

这种交叉对角线乘积的方法是计算三阶行列式的一种简洁方式。但遗憾的是，它难以应用于高于三阶行列式的计算。对这种高于三阶的行列式，我们必须借助于所谓的“拉普拉斯展开”。

## 用拉普拉斯展开计算 *n* 阶行列式

首先，我们以三阶行列式为例解释拉普拉斯展开过程。回到(5.6)式的第一行，我们看到 $|A|$ 的值也可以看作三项之和，其中每一项为第一行的元素与一特定的二阶行列式之积。通过某些

低阶行列式计算 $|A|$ 的过程说明了行列式的拉普拉斯展开。

(5.6)中的三个二阶行列式不是任意确定的，而是由特定的规律设定的。第一个二阶行列式 $\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$ 是一个通过删除 $|A|$ 的第一行和第一列而得到的子行列式。它被称作元素 $a_{11}$(位于删除的行与列的交叉点)的子式，并以 $|M_{11}|$ 表示。一般而言，符号 $|M_{ij}|$ 可以用于表示删除给定行列式的第 $i$ 行第 $j$ 列而得到的子式。因子式本身是一个行列式，它具有一个值。读者可以自己验证，(5.6)中其它两个二阶行列式分别是子式 $|M_{12}|$ 和 $|M_{13}|$，即

96

$$|M_{11}| \equiv \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} \quad |M_{12}| \equiv \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} \quad |M_{13}| \equiv \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

与子式密切相关的一个概念是余子式。余子式以 $|C_{ij}|$ 表示，是规定了代数符号的子式。[①] 其符号确定的规则如下：若子式 $|M_{ij}|$ 中下标 $i$ 和 $j$ 之和为偶数，则余子式与子式符号相同，即 $|C_{ij}| = |M_{ij}|$。若为奇数，则余子式符号与子式相反，即 $|C_{ij}| = -|M_{ij}|$。简言之，我们有：

$$|C_{ij}| = (-1)^{i+j}|M_{ij}|$$

这里很明显，当且仅当 $(i+j)$ 为偶数，表达式 $(-1)^{i+j}$ 为正。余子式具有特定的符号是极其重要的，应牢记在心。

**例 4**

---

① 很多作者使用符号 $M_{ij}$ 和 $C_{ij}$(没有两条纵线)来表示子式和余子式。我们加上两条短纵线是为了直观地强调这一事实：子式和余子式具有行列式性质，因此具有标量值。

行列式$\begin{vmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{vmatrix}$中，元素 8 的子式为

$$|M_{12}| = \begin{vmatrix} 6 & 4 \\ 3 & 1 \end{vmatrix} = -6$$

但同一元素的余子式为

$$|C_{12}| = -|M_{12}| = 6$$

因为 $i+j=1+2=3$ 为奇数。类似地，元素 4 的余子式为

$$|C_{23}| = -|M_{23}| = -\begin{vmatrix} 9 & 8 \\ 3 & 2 \end{vmatrix} = 6$$

运用这些新的概念，我们可以将三阶行列式表示成：

$$|A| = a_{11}|M_{11}| - a_{12}|M_{12}| + a_{13}|M_{13}|$$

$$= a_{11}|C_{11}| + a_{12}|C_{12}| + a_{13}|C_{13}| = \sum_{j=1}^{3} a_{1j}|C_{1j}| \tag{5.7}$$

即表示成三项之和，其中每一项是第一行的元素与其对应的余子式的积，注意在(5.7)中，$a_{12}|M_{12}|$ 与 $a_{12}|C_{12}|$ 的符号不同。这是因为 $1+2$ 为奇数。

三阶行列式的拉普拉斯展开可以使三阶行列式的计算问题简化为某些二阶行列式的计算问题。类似的简化在高阶行列式的拉 97 普拉斯展开中也可以实现。例如，在四阶行列式 $|B|$ 中，第一行包含四个元素 $b_{11},\cdots,b_{14}$，因此，根据(5.7)式的精神，我们可以写出

$$|B| = \sum_{j=1}^{4} b_{1j}|C_{1j}|$$

其中余子式 $|C_{1j}|$ 是 3 阶的。每个三阶余子式则可按(5.6)式计算。一般而言，$n$ 阶行列式的拉普拉斯展开可把问题简化为计算 $n$ 个 $(n-1)$ 阶余子式；重复运用这种展开过程，将使行列式的阶数越来越低，最终获得如(5.5)所定义的基本的二阶行列式。则初始行列式的值就很容易计算了。

尽管拉普拉斯展开过程是按照第一行元素的余子式来表达的，但也可以用任意一行，或者任意一列去展开一个行列式。例如，若一个三阶行列式 $|A|$ 的第 1 列包含元素 $a_{11}$，$a_{21}$，$a_{31}$，这些元素的余子式展开将会产生 $|A|$ 值：

$$|A| = a_{11}|C_{11}| + a_{21}|C_{21}| + a_{31}|C_{31}| = \sum_{i=1}^{3} a_{i1}|C_{i1}|$$

**例 5** 已知 $A = \begin{vmatrix} 5 & 6 & 1 \\ 2 & 3 & 0 \\ 7 & -3 & 0 \end{vmatrix}$，由第一行展开产出结果

$$|A| = 5\begin{vmatrix} 3 & 0 \\ -3 & 0 \end{vmatrix} - 6\begin{vmatrix} 2 & 0 \\ 7 & 0 \end{vmatrix} + \begin{vmatrix} 2 & 3 \\ 7 & -3 \end{vmatrix} = 0 + 0 - 27 = -27$$

但由第一列展开也得到一致的答案

$$|A| = 5\begin{vmatrix} 3 & 0 \\ -3 & 0 \end{vmatrix} - 2\begin{vmatrix} 6 & 1 \\ -3 & 0 \end{vmatrix} + 7\begin{vmatrix} 6 & 1 \\ 3 & 0 \end{vmatrix} = 0 - 6 - 21 = -27$$

迄今为止我们仅涉及数字运算，这一事实使我们有机会选择某些“容易”的行或列去展开。具有最多 0 或 1 的行或列最易展开，因为 0 乘以其余子式等于 0，所以该项可以舍弃，1 乘以余子式等于余子式自身，所以至少可以省略一个乘积步骤。在例 5 中，展开行列式的最容易的方式是用第三列展开，该列包含元素 1,0,0。

因而我们可以这样计算：

$$|A| = 1\begin{vmatrix}2 & 3\\ 7 & -3\end{vmatrix} - 0 + 0 = -27$$

98　　总之，$n$ 阶行列式 $|A|$ 的值可以通过**任意行或任意列的拉普拉斯展开**来计算：

$$\begin{aligned} |A| &= \sum_{j=1}^{n} a_{ij}|C_{ij}| \quad [\text{由第 } i \text{ 行展开}] \\ &= \sum_{i=1}^{n} a_{ij}|C_{ij}| \quad [\text{由第 } j \text{ 列展开}] \end{aligned} \tag{5.8}$$

## 练习 5.2

---

1　计算下列行列式：

(a) $\begin{vmatrix}8 & 1 & 3\\ 4 & 0 & 1\\ 6 & 0 & 3\end{vmatrix}$　(c) $\begin{vmatrix}4 & 0 & 2\\ 6 & 0 & 3\\ 8 & 2 & 3\end{vmatrix}$　(e) $\begin{vmatrix}a & b & c\\ b & c & a\\ c & a & b\end{vmatrix}$

(b) $\begin{vmatrix}1 & 2 & 3\\ 4 & 7 & 5\\ 3 & 6 & 9\end{vmatrix}$　(d) $\begin{vmatrix}1 & 1 & 4\\ 8 & 11 & -2\\ 0 & 4 & 7\end{vmatrix}$　(f) $\begin{vmatrix}x & 5 & 0\\ 3 & y & 2\\ 9 & -1 & 8\end{vmatrix}$

2　为求得下列余子式：

$|C_{13}|$，$|C_{23}|$，$|C_{33}|$，$|C_{41}|$ 及 $|C_{34}|$

请确定相应子式前的符号。

3　已知 $\begin{vmatrix}a & b & c\\ d & e & f\\ g & h & i\end{vmatrix}$，求元素 $a$，$b$，$f$ 的子式和余子式。

4　计算下列行列式：

$$(a)\begin{vmatrix}1 & 2 & 0 & 9\\2 & 3 & 4 & 6\\1 & 6 & 0 & -1\\0 & -5 & 0 & 8\end{vmatrix}\quad(b)\begin{vmatrix}2 & 7 & 0 & 1\\5 & 6 & 4 & 8\\0 & 0 & 9 & 0\\1 & -3 & 1 & 4\end{vmatrix}$$

5　在上题的第一个行列式中,求元素9的余子式的值。

---

## 5.3　行列式的基本性质

现在,我们可以讨论行列式的某些性质,从而使我们可以"发现"方阵行线性相关与矩阵行列式等于零之间的关系。

这里讨论五个基本性质。尽管我们在多数情况下将以二阶行列式为例来说明,但这些性质对各阶行列式均适用。

**性质 I** 行与列互换不影响行列式的值。换言之,矩阵 $A$ 的行列式与其转置矩阵 $A'$ 的行列式具有相同的值,即 $|A| = |A'|$。

**例 1**　$\begin{vmatrix}4 & 3\\5 & 6\end{vmatrix} = \begin{vmatrix}4 & 5\\3 & 6\end{vmatrix} = 9$

**例 2**　$\begin{vmatrix}a & b\\c & d\end{vmatrix} = \begin{vmatrix}a & c\\b & d\end{vmatrix} = ad - bc$

**性质 II** 任意两行或两列互换将改变行列式符号,但数值不变。

**例 3**　$\begin{vmatrix}a & b\\c & d\end{vmatrix} = ad - bc$,但两行互换有

$$\begin{vmatrix}c & d\\a & b\end{vmatrix} = cb - ad = -(ad - bc)$$

**例 4** $\begin{vmatrix} 0 & 1 & 3 \\ 2 & 5 & 7 \\ 3 & 0 & 1 \end{vmatrix} = -26$，但第一列与第三列互换得到

$$\begin{vmatrix} 3 & 1 & 0 \\ 7 & 5 & 2 \\ 1 & 0 & 3 \end{vmatrix} = 26.$$

99 **性质 III** 以一标量 $k$ 乘行列式任意一行(或任意一列)，行列式的值将发生 $k$ 倍变化。

**例 5** 以 $k$ 乘例 3 中行列式的第一行，我们得到

$$\begin{vmatrix} ka & kb \\ c & d \end{vmatrix} = kad - kbc = k(ad - bc) = k\begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

区分两个表达式 $kA$ 与 $k|A|$ 之间的差别是重要的。在以标量 $k$ 乘矩阵 A 时，A 中所有元素均乘以 $k$。但若我们从左到右读本例中的方程，很明显，以 $k$ 乘以行列式 $|A|$，仅有一行或列被 $k$ 乘。因此，这一方程实际上给出了一个分解行列式的法则：当任意一行或列含有公约数时，就可以将其提到行列式外面。

**例 6** 依次提取第一列和第一行的公约数，我们有

$$\begin{vmatrix} 15a & 7b \\ 12c & 2d \end{vmatrix} = 3\begin{vmatrix} 5a & 7b \\ 4c & 2d \end{vmatrix} = 3(2)\begin{vmatrix} 5a & 7b \\ 2c & d \end{vmatrix} = 6(5ad - 14bc)$$

100 当然，直接计算原行列式也会产生同样结果。

相反，提取矩阵的公约数要求矩阵中所有元素具有公约数，如下式：

$$\begin{bmatrix} ka & kb \\ kc & kd \end{bmatrix} = k\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

**性质 IV** 行列式的某一行加上(减去)另一行的倍数，其值不

变。将上述表述中的“行”变为“列”,也同样成立。

**例 7** 将例 3 中的行列式的第一行乘以 $k$,并加到第二行上,我们以得到原行列式而结束计算:

$$\begin{vmatrix} a & b \\ c+ka & d+kb \end{vmatrix} a(d+kb)-b(c+ka)=ad-bc=\begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

**性质 V** 若行列式的一行(或列)为另一行(或列)的倍数,则行列式的值为零。作为这种情况的一个特例,当行列式两行(或两列)相同时,行列式将变成零。

**例 8** $\begin{vmatrix} 2a & 2b \\ a & b \end{vmatrix}=2ab-2ab=0 \quad \begin{vmatrix} c & c \\ d & d \end{vmatrix}=cd-cd=0$

行列式“变成零”的另外一个例子在练习 5.2—1 中可找到。

这个重要性质实际是性质 IV 的逻辑结果。为理解这一点,我们将性质 IV 应用于例 8 中的两个行列式并观察其结果。对于第一个行列式,从第一行减去第二行的二倍;对于第二个行列式,从第一列减去第二列的二倍。因为这些运算不改变行列式的值,我们可以写成

$$\begin{vmatrix} 2a & 2b \\ a & b \end{vmatrix}=\begin{vmatrix} 0 & 0 \\ a & b \end{vmatrix} \quad \begin{vmatrix} c & c \\ d & d \end{vmatrix}=\begin{vmatrix} 0 & c \\ 0 & d \end{vmatrix}$$

现在新的(化简的)行列式分别有一行和一列为零,因此在这两个例子中,其拉普拉斯展开一定为零。一般而言,当行列式的一行(列)为另一行(列)的倍数时,应用性质 IV 总可以将该行(列)的所有元素化简为零,因此,性质 V 成立。

刚才讨论的基本性质有几方面的用途。首先,它有助于极大地简化行列式的计算。例如,通过从一行(或列)中减去另一行(或列)的倍数,行列式中的元素可以化简成比较小和简单的数。若可

能的话，提取公约数也具有同样的效果。若我们真能应用这些性质将某些行或列转化成尽可能包含更多的 0 和 1 的形式，则行列式的拉普拉斯展开将会是非常轻松的工作。

101

### 非奇异性行列式的判别标准

我们现在主要关心的是行线性相关与行列式为零之间的联系。为此，可以引用性质 V。考察方程组 $Ax=d$：

$$\begin{bmatrix} 3 & 4 & 2 \\ 15 & 20 & 10 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$$

当且仅当系数矩阵 $A$ 的行线性无关，方程组具有唯一解，使得 $A$ 为非奇异矩阵。但矩阵第二行为第一行的 5 倍，行实际上线性相关，因此没有唯一解。这个行相关是通过直观观察确定的，但运用性质 V，通过 $|A|=0$ 这一事实，也可以发现这一点。

当然，矩阵行相关还可以假设为更复杂、隐蔽的形式。例如在下面矩阵中

$$B = \begin{bmatrix} 4 & 1 & 2 \\ 5 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} v'_1 \\ v'_2 \\ v'_3 \end{bmatrix}$$

就存在行相关，因为 $2v_1' - v_2' - 3v_3 = 0$；但这一事实难以通过直观观测发现。由性质 V 可知行列式为零，即 $|B|=0$，因为 3 倍的 $v_3'$ 加上 $v_2'$ 减去 2 倍的 $v_1'$，可使第二行化简为零向量。一般而言，任何类型行线性相关均可由一个等于零的行列式来反映，性质 V 的美妙之处也正在这里。相反，若行线性无关，则行列式一

定有非零值。

在上面我们已把矩阵非奇异性与行线性无关基本联系起来了。但有时我们需要这一结论：对于一个方阵 $A$，行线性无关$\Leftrightarrow$列线性无关，现在我们来对其加以证明：

根据性质 I，我们知道 $|A| = |A'|$。因为 $A$ 中行线性无关$\Leftrightarrow |A| \neq 0$，所以也可以说 $A$ 中行线性无关$\Leftrightarrow |A'| \neq 0$。但 $|A'| \neq 0 \Leftrightarrow$转置矩阵 $A'$中行线性无关$\Leftrightarrow A$ 中列线性无关（据定义 $A'$的行为 $A$ 的列）因而 $A$ 中行线性无关$\Leftrightarrow A$ 中列线性无关。

现在，我们把关于非奇异性检验的讨论总结一下。给定线性 102 方程组 $Ax = d$，其中 $A$ 为一 $n \times n$ 系数矩阵，

$|A| \neq 0 \Leftrightarrow$矩阵 $A$ 中行（列）线性无关

$\Leftrightarrow A$ 为非奇异矩阵

$\Leftrightarrow A^{-1}$存在

$\Leftrightarrow$一个惟一解 $\bar{x} = A^{-1}d$ 存在

因而，系数矩阵行列式的值 $|A|$，提供了一个检验矩阵 $A$ 非奇异性和方程组 $Ax = d$ 存在惟一解的非常方便的判别标准。但要注意，行列式判别标准并未提到解值的代数符号，尽管我们可以肯定当 $|A| \neq 0$ 时，存在惟一解，但有时我们也可能得到经济上不允许的负的解值。

**例 9** 下列方程组有惟一解吗？

$$
\begin{aligned}
7x_1 - 3x_2 - 3x_3 &= 7 \\
2x_1 + 4x_2 + x_3 &= 0 \\
-2x_2 - x_3 &= 2
\end{aligned}
$$

行列式 $|A|$ 为

$$\begin{vmatrix} 7 & -3 & -3 \\ 2 & 4 & 1 \\ 0 & -2 & -1 \end{vmatrix} = -8 \neq 0$$

因而确实存在唯一解。…

## 重新定义矩阵的秩

前面我们把矩阵 $A$ 的秩定义为 $A$ 中线性无关的行的最大数量。根据行线性不相关与其行列式不为零之间的关系，我们可以103将 $m \times n$ 矩阵的秩重新定义为由该矩阵的行和列所构造的不为零的行列式的最大阶数。任何矩阵的秩都是一个惟一的值。

显然，$m \times n$ 矩阵的最大秩为 $m$ 与 $n$ 中较小的一个。因为行列式仅对方阵有定义，由矩阵的维数，比如 $3 \times 5$ 矩阵，其行列式最大可能的阶数为 3。这一事实用符号表示如下：

$$r(A) \leqslant \min\{m, n\}$$

读作："$A$ 的秩小于或等于由 $m$ 和 $n$ 组成的集合中最小的一个数"。$n \times n$ 非奇异矩阵的秩必为 $n$，在此情况下，我们可以写成：$r(A) = n$。

有时，有人可能会对两矩阵积的秩感兴趣。在这种情况下，可运用下列法则：

$$r(AB) \leqslant \min\{r(A), r(B)\}$$

## 练习 5.3

---

1 用行列式 $\begin{vmatrix} 2 & 0 & -1 \\ 1 & 1 & 7 \\ 3 & 3 & 9 \end{vmatrix}$ 验证行列式的前四个性质。

2 证明:当 $n$ 阶行列式 $|A|$ 的所有元素被数 $k$ 乘时,结果将为 $k^n|A|$。

3 行列式的哪一性质使我们可以作下列运算?

(a) $\begin{vmatrix} 9 & 18 \\ 27 & 56 \end{vmatrix} = \begin{vmatrix} 9 & 18 \\ 0 & 2 \end{vmatrix}$ (b) $\begin{vmatrix} 9 & 27 \\ 4 & 2 \end{vmatrix} = 18\begin{vmatrix} 1 & 3 \\ 2 & 1 \end{vmatrix}$

4 检验下列行列式是否为非奇异矩阵:

(a) $\begin{bmatrix} 4 & 0 & 1 \\ 19 & 1 & 3 \\ 5 & 4 & 7 \end{bmatrix}$ (c) $\begin{bmatrix} 7 & -1 & 0 \\ 1 & 1 & 4 \\ 13 & -3 & -4 \end{bmatrix}$

(b) $\begin{bmatrix} 4 & -2 & 1 \\ -5 & 6 & 0 \\ 7 & 0 & 3 \end{bmatrix}$ (d) $\begin{bmatrix} 7 & 9 & 5 \\ 3 & 0 & 1 \\ 10 & 8 & 6 \end{bmatrix}$

5 对上题中每一矩阵的秩可以得出何结论?

6 下面每组 3-向量能生成 3-空间吗?为什么能?或者为什么不能?

(a) $[1\ \ 2\ \ 0]\quad[2\ \ 3\ \ 1]\quad[3\ \ 4\ \ 2]$

(b) $[8\ \ 1\ \ 3]\quad[1\ \ 2\ \ 8]\quad[-7\ \ 1\ \ 5]$

7 用 $Ax=d$ 的形式改写(3.23)简单国民收入模型,然后检验系数矩阵 $A$ 是否为非奇异矩阵。(把 $Y$ 作为向量 $x$ 中的第一个变量。)

---

## 5.4 求逆矩阵

若线性方程组 $Ax=d$ 中的矩阵 $A$ 为非奇异矩阵,则 $A^{-1}$ 存在,且方程组的解为 $\bar{x}=A^{-1}d$。我们已知道,通过判别标准 $|A|\neq 0$,可以检验 $A$ 的非奇异性。下一个问题是:若 $A$ 确实通过了非奇异性检验,我们如何求逆 $A^{-1}$。

### 按异行余子式展开行列式

在回答这个问题之前,我们先讨论行列式的另一个重要性质:

**性质 VI** 按异行余子式（未按“正确”的行或列展开的余子式）展开的行列式，其值恒为零。

**例 1** 若我们用第一行的元素和第二行中对应元素的余子式

104 来展开行列式 $\begin{vmatrix} 4 & 1 & 2 \\ 5 & 2 & 1 \\ 1 & 0 & 3 \end{vmatrix}$

第二行元素的余子式为

$$|C_{21}| = -\begin{vmatrix} 1 & 2 \\ 0 & 3 \end{vmatrix} = -3 \qquad |C_{22}| = \begin{vmatrix} 4 & 2 \\ 1 & 3 \end{vmatrix} = 10$$

$$|C_{23}| = -\begin{vmatrix} 4 & 1 \\ 1 & 0 \end{vmatrix} = 1$$

我们得到

$$a_{11}|C_{21}| + a_{12}|C_{22}| + a_{13}|C_{23}| = 4(-3) + 1(10) + 2(1) = 0.$$

更一般地，把例 1 所述的按异行余子式展开行列式的方法应用于行列式 $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$ 得到的如下乘积之和也为零：

$$\begin{aligned}
\sum_{j=1}^{3} a_{1j}|C_{2j}| &= a_{11}|C_{21}| + a_{12}|C_{22}| + a_{13}|C_{23}| \\
&= -a_{11}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{12}\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - a_{13}\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}] \\
&= -a_{11}a_{12}a_{33} + a_{11}a_{13}a_{32} + a_{11}a_{12}a_{33} - a_{12}a_{13}a_{31} \\
&\quad - a_{11}a_{13}a_{32} + a_{12}a_{13}a_{31} = 0
\end{aligned} \tag{5.9}$$

产生这一结果的原因在于这一事实：(5.9) 中乘积之和可以视为另

一行列式

$$|A^*| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

按第二行正规展开的结果，$|A^*|$ 与 $|A|$ 的区别仅在于第二行，$|A^*|$ 的前两行是相同的。作为一个练习，读者可以自己写出 $|A^*|$ 第二行的余子式，并验证它们确实是(5.9)中出现的余子式，并且具有正确的符号。因为 $|A^*|$ 的前两行相同，所以 $|A^*|=0$，因而(5.9)所示的按异行余子式展开的行列式必须也为零。

性质 VI 对各阶行列式均成立；并且当行列式按任意行或列的异行余子式展开时也可应用此性质，因而我们一般可断言，对于 $n$ 阶行列式，下列结论成立：

$$\sum_{j=1}^{n} a_{ij}|C_{i'j}| = 0 \qquad (i \neq i')$$

［按第 $i$ 行和第 $i'$ 行的余子式展开］

$$\sum_{i=1}^{n} a_{ij}|C_{ij'}| = 0 \qquad (j \neq j') \qquad (5.10)$$

［按第 $j$ 列和第 $j'$ 列的余子式展开］

仔细比较(5.10)与(5.8)。(5.8)式(正规的拉普拉斯展开)在和中的每一乘积项中 $a_{ij}$ 和 $|C_{ij}|$ 的下标必相同；而在像(5.10)的异行余子式的展开中，两个下标中的一个(一选定的 $i'$ 或 $j'$ 的值)必定是不相同的。

## 矩阵求逆

在(5.10)中归纳的性质 VI,对于发展一种矩阵求逆的方法,即求出逆矩阵的方法,具有直接的帮助。

假设给定 $n \times n$ 非奇异矩阵A

$$\underset{(n\times n)}{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \quad (|A| \neq 0) \tag{5.11}$$

因 $A$ 中每一元素具有余子式 $|C_{ij}|$,有可能通过以余子式 $|C_{ij}|$ 置换(5.11)中每一元素 $a_{ij}$ 而形成一个余子式矩阵。这样的余子式矩阵,以 $C=[|C_{ij}|]$表示,必为 $n \times n$ 矩阵。但基于现在之目的,我们对 $C$ 的转置更感兴趣。转置 $C'$一般称作 $A$ 的**伴随矩阵**,以符号 adj $A$ 表示。伴随矩阵的形式如下:

$$\underset{(n\times n)}{C'} \equiv \text{adj } A \equiv \begin{bmatrix} |C_{11}| & |C_{21}| & \cdots & |C_{n1}| \\ |C_{12}| & |C_{22}| & \cdots & |C_{n2}| \\ \cdots & \cdots & \cdots & \cdots \\ |C_{1n}| & |C_{2n}| & \cdots & |C_{nn}| \end{bmatrix} \tag{5.12}$$

矩阵 $A$ 和$C'$可相乘,其积 $AC'$为另一个 $n \times n$ 矩阵,其中的每一元素为乘积和。通过利用拉普拉斯展开公式和行列式性质

VI,乘积 $AC'$ 可以表示如下：

$$
\underset{(n\times n)}{AC'} = \begin{bmatrix} \sum_{j=1}^{n} a_{1j}|C_{1j}| & \sum_{j=1}^{n} a_{1j}|C_{2j}| & \cdots & \sum_{j=1}^{n} a_{1j}|C_{nj}| \\ \sum_{j=1}^{n} a_{2j}|C_{1j}| & \sum_{j=1}^{n} a_{2j}|C_{2j}| & \cdots & \sum_{j=1}^{n} a_{2j}|C_{nj}| \\ \vdots & \vdots & & \vdots \\ \sum_{j=1}^{n} a_{nj}|C_{1j}| & \sum_{j=1}^{n} a_{nj}|C_{2j}| & \cdots & \sum_{j=1}^{n} a_{nj}|C_{nj}| \end{bmatrix}
$$

$$
= \begin{bmatrix} |A| & 0 & \cdots & 0 \\ 0 & |A| & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & |A| \end{bmatrix} \text{[由(5.8)和(5.10)]}
$$

$$
= |A| \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = |A| I_n \text{[因式分解]}
$$

因为行列式 $|A|$ 为一非零标量，所以可以将方程 $AC' = |A| I$ 两边同除以 $|A|$，结果为：106

$$
\frac{AC'}{|A|} = I \text{ 或 } A\frac{C'}{|A|} = I
$$

以 $A^{-1}$ 同时左乘后一方程两边，并利用结果 $A^{-1}A = I$，我们得到

$\frac{C'}{|A|} = A^{-1}$，或者

$$A^{-1} = \frac{1}{|A|}\operatorname{adj} A \quad [\text{由}(5.12)] \tag{5.13}$$

现在,我们已找到求矩阵 $A$ 的逆的方法了。

求方阵 $A$ 的逆矩阵的一般步骤可概括如下:(1) 求 $|A|$〔当且仅当 $|A| \neq 0$,我们才可进行下一步骤,因若 $|A| = 0$,(5.13)中的逆无定义〕;(2) 求 $A$ 所有元素的余子式,将其排成余子式矩阵 $C = [\,|C_{ij}|\,]$;(3) 求 $C$ 的转置以得到 adj $A$;(4) 以行列式 $|A|$ 除 adj $A$,其结果便是所要求的逆矩阵 $A^{-1}$

**例 2** 求 $A = \begin{bmatrix} 3 & 2 \\ 1 & 0 \end{bmatrix}$ 的逆矩阵。因 $|A| = -2 \neq 0$,逆矩阵 $A^{-1}$存在。在本例中每一元素的余子式为 $1 \times 1$ 行列式,它简单定义为行列式本身的标量元素(即 $|a_{ij}| \equiv a_{ij}$)。因此我们有:

$$C = \begin{bmatrix} |C_{11}| & |C_{12}| \\ |C_{21}| & |C_{22}| \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -2 & 3 \end{bmatrix}$$

注意,1 和 2 前面有负号,正如余子式所要求的那样。转置余子式矩阵得到

$$\operatorname{adj} A = \begin{bmatrix} 0 & -2 \\ -1 & 3 \end{bmatrix}$$

所以逆矩阵 $A^{-1}$可以写成

$$A^{-1} = \frac{1}{|A|}\operatorname{adj} A = -\frac{1}{2}\begin{bmatrix} 0 & -2 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & -\frac{3}{2} \end{bmatrix}$$

**例 3** 求矩阵 $B = \begin{bmatrix} 4 & 1 & -1 \\ 0 & 3 & 2 \\ 3 & 0 & 7 \end{bmatrix}$ 的逆矩阵。

107 因 $|B| = 99 \neq 0$,所以逆矩阵 $B^{-1}$存在。余子式矩阵为:

$$\begin{bmatrix} \begin{vmatrix} 3 & 2 \\ 0 & 7 \end{vmatrix} & -\begin{vmatrix} 0 & 2 \\ 3 & 7 \end{vmatrix} & \begin{vmatrix} 0 & 3 \\ 3 & 0 \end{vmatrix} \\ -\begin{vmatrix} 1 & -1 \\ 0 & 7 \end{vmatrix} & \begin{vmatrix} 4 & -1 \\ 3 & 7 \end{vmatrix} & -\begin{vmatrix} 4 & 1 \\ 3 & 0 \end{vmatrix} \\ \begin{vmatrix} 1 & -1 \\ 3 & 2 \end{vmatrix} & -\begin{vmatrix} 4 & -1 \\ 0 & 2 \end{vmatrix} & \begin{vmatrix} 4 & 1 \\ 0 & 3 \end{vmatrix} \end{bmatrix} = \begin{bmatrix} 21 & 6 & -9 \\ -7 & 31 & 3 \\ 5 & -8 & 12 \end{bmatrix}$$

因而

$$\operatorname{adj} B = \begin{bmatrix} 21 & -7 & 5 \\ 6 & 31 & -8 \\ -9 & 3 & 12 \end{bmatrix}$$

且所求逆矩阵为

$$B^{-1} = \frac{1}{|B|}\operatorname{adj} B = \frac{1}{99}\begin{bmatrix} 21 & -7 & 5 \\ 6 & 31 & -8 \\ -9 & 3 & 12 \end{bmatrix}$$

读者可以检验,以上两例的结果分别满足

$$AA^{-1} = A^{-1}A = I \text{ 和 } BB^{-1} = B^{-1}B = I。$$

## 练习 5.4

108

---

1 假定我们按第三列和第二列元素的余子式展开一个四阶行列式,如何用$\sum$符号写出所得到的乘积之和?若我们按第二行和第四行元素的余子式展开此行列式,以$\sum$符号表示的乘积和将会如何呢?

2 求下列矩阵的逆。

(a) $A = \begin{bmatrix} 5 & 2 \\ 0 & 1 \end{bmatrix}$　　(c) $C = \begin{bmatrix} 7 & 7 \\ 3 & -1 \end{bmatrix}$

(b) $B = \begin{bmatrix} 1 & 0 \\ 9 & 2 \end{bmatrix}$　　(d) $D = \begin{bmatrix} 7 & 6 \\ 0 & 3 \end{bmatrix}$

3 (*a*) 利用上一题的答案，推出一个求一已知 $2\times 2$ 矩阵 $A$ 的伴随矩阵的两步法则：第一步，指明为求出 adj $A$ 的对角元素，应如何处理 $A$ 的两个对角元素；第二步，如何处理 $A$ 的两个非对角元素。[注意，此法则仅适于 $2\times 2$ 矩阵。]

(*b*) 在上述两步基础上，加上第三步，求出 $2\times 2$ 逆矩阵 $A^{-1}$。

4 求下列矩阵的逆矩阵：

(*a*) $E=\begin{bmatrix}4 & -2 & 1\\ 7 & 3 & 3\\ 2 & 0 & 1\end{bmatrix}$ (*c*) $G=\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}$

(*b*) $F=\begin{bmatrix}1 & -1 & 2\\ 1 & 0 & 3\\ 4 & 0 & 2\end{bmatrix}$ (*d*) $H=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$

5 一个矩阵可能是其自身的逆矩阵吗？

---

# 5.5 克莱姆法则

刚才讨论的矩阵求逆的方法，使我们得以推导出一种方便实用的解线性方程组的方法，这种方法称作**克莱姆法则**。

## 克莱姆法则的推导

给定方程组 $Ax=d$，其中 $A$ 为 $n\times n$ 矩阵。若 $A$ 为非奇异矩阵，方程组的解可以写成

$$\bar{x}=A^{-1}d=\frac{1}{|A|}(\text{adj } A)d \qquad [\text{由}(5.13)]$$

根据(5.12)，这意味着

$$\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_n \end{bmatrix} = \frac{1}{|A|} \begin{bmatrix} |C_{11}| & |C_{21}| & \cdots & |C_{n1}| \\ |C_{12}| & |C_{22}| & \cdots & |C_{n2}| \\ \cdots & \cdots & \cdots & \cdots \\ |C_{1n}| & |C_{2n}| & \cdots & |C_{nn}| \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$$

$$= \frac{1}{|A|} \begin{bmatrix} d_1|C_{11}| + d_2|C_{21}| + \cdots + d_n|C_{n1}| \\ d_1|C_{12}| + d_2|C_{22}| + \cdots + d_n|C_{n2}| \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ d_1|C_{1n}| + d_2|C_{2n}| + \cdots + d_n|C_{nn}| \end{bmatrix}$$

$$= \frac{1}{|A|} \begin{bmatrix} \sum_{i=1}^{n} d_i |C_{i1}| \\ \sum_{i=1}^{n} d_i |C_{i2}| \\ \vdots \\ \sum_{i=1}^{n} d_i |C_{in}| \end{bmatrix}$$

使方程两边对应元素相等,我们得到解值

$$\bar{x}_1 = \frac{1}{|A|}\sum_{i=1}^{n} d_i |C_{i1}| \qquad \bar{x}_2 = \frac{1}{|A|}\sum_{i=1}^{n} d_i |C_{i2}| \quad (\text{等}) \tag{5.14}$$

(5.14)中的$\sum$项看起来有些陌生。它们表示什么呢?由(5.8)我们知道,行列式$|A|$按第一列的拉普拉斯展开可以表示成$\sum_{i=1}^{n} a_{i1}|C_{i1}|$。若我们以列向量$d$置换$|A|$中的第一列,

但保持 $|A|$ 中的其它各列不变，则得到一个新的行列式，称其为 $|A_1|$，下标 1 标明第一列已被 $d$ 置换。按第一列（$d$ 列）展开 $|A_1|$ 得到表达式 $\sum_{i=1}^{n} d_i |C_{i1}|$，因为现在元素 $d_i$ 已代替了元素 $a_{i1}$。回到（5.14），我们便看到

109

$$\bar{x}_1 = \frac{1}{|A|}|A_1|$$

类似地，若我们以列向量 $d$ 置换 $|A|$ 中的第二列，而保持其它各列不变，按第二列（$d$ 列）展开新行列式 $|A_2|$，得到表达式 $\sum_{i=1}^{n} d_i |C_{i2}|$。被 $|A|$ 除，便得到 $\bar{x}_2$ 的解值；如此等等。

这个过程可归纳如下：要求出第 $j$ 个变量 $\bar{x}_j$ 的解值，我们仅需以常数项 $d_1, \cdots, d_n$ 代替行列式 $|A|$ 中的第 $j$ 列，得到新行列式 $|A_j|$，然后以原行列式 $|A|$ 除以 $|A_j|$。这样，方程组 $Ax = d$ 的解可以表示成

$$\bar{x}_j = \frac{|A_j|}{|A|} = \frac{1}{|A|}\begin{vmatrix} a_{11} & a_{12} & \cdots & d_1 & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & d_2 & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & d_n & \cdots & a_{nn} \end{vmatrix} \tag{5.15}$$

（由 $d$ 代替第 $j$ 列）

（5.15）中的结果便是克莱姆法则的表述。注意，矩阵求逆的方法一次得到**所有**内生变量的解值（$\bar{x}$ 是一个向量），但克莱姆法则一次只得到一个内生变量的解值（$\bar{x}_j$ 是一个标量）。

**例 1** 求下列方程组的解

$$5x_1 + 3x_2 = 30$$

$$6x_1 - 2x_2 = 8$$

由系数和常数项可得下列矩阵

$$|A| = \begin{vmatrix} 5 & 3 \\ 6 & -2 \end{vmatrix} = -28 \qquad |A_1| = \begin{vmatrix} 30 & 3 \\ 8 & -2 \end{vmatrix} = -84$$

$$|A_2| = \begin{vmatrix} 5 & 30 \\ 6 & 8 \end{vmatrix} = -140$$

因此，由(5.15)式，我们马上得到

$$\bar{x}_1 = \frac{|A_1|}{|A|} = \frac{-84}{-28} = 3 \quad 和 \quad \bar{x}_2 = \frac{|A_2|}{|A|} = \frac{-140}{-28} = 5$$

**例 2** 求下列方程组的解

$$7x_1 - x_2 - x_3 = 0$$

$$10x_1 - 2x_2 + x_3 = 8$$

110

$$6x_1 + 3x_2 - 2x_3 = 7$$

可求得相应的行列式 $|A|$ 和 $|A_j|$ 为

$$|A| = \begin{vmatrix} 7 & -1 & -1 \\ 10 & -2 & 1 \\ 6 & 3 & -2 \end{vmatrix} = -61 \qquad |A_1| = \begin{vmatrix} 0 & -1 & -1 \\ 8 & -2 & 1 \\ 7 & 3 & -2 \end{vmatrix} = -61$$

$$|A_2| = \begin{vmatrix} 7 & 0 & -1 \\ 10 & 8 & 1 \\ 6 & 7 & -2 \end{vmatrix} = -183 \qquad |A_3| = \begin{vmatrix} 7 & -1 & 0 \\ 10 & -2 & 8 \\ 6 & 3 & 7 \end{vmatrix} = -244$$

因而变量解值为

$$\bar{x}_1 = \frac{|A_1|}{|A|} = \frac{-61}{-61} = 1 \qquad \bar{x}_2 = \frac{|A_2|}{|A|} = \frac{-183}{-61} = 3$$

$$\bar{x}_3 = \frac{|A_3|}{|A|} = \frac{-244}{-61} = 4$$

注意，在上面每一例中，我们求得 $|A| \neq 0$。这是应用克莱姆法则的必要条件，因为它也是逆矩阵 $A^{-1}$存在的必要条件。克莱姆法则在计算时绕过了矩阵求逆的过程，但它毕竟是建立在矩阵求逆概念基础上的。

## 对齐次方程组的阐释

上面所考察的方程组 $Ax = d$ 在向量 $d$ 中可具有任意常数。但若 $d = 0$，即若 $d_1 = d_2 = \cdots = d_n = 0$，方程组将变成

$$Ax = 0$$

其中 0 为零向量。这种特殊情况被称作**齐次方程组**。[①]

若 $A$ 为非奇异矩阵，齐次方程组仅能得到一个“平凡解”，即 $\bar{x}_1 = \bar{x}_2 = \cdots = \bar{x}_n = 0$。这一结论是从下述事实推导出来的：解 $\bar{x} = A^{-1}d$在这种情况下会变成

$$\underset{(n\times 1)}{\bar{x}} = \underset{(n\times n)}{A^{-1}} \underset{(n\times 1)}{0} = \underset{(n\times 1)}{0}$$

另外，还可以从克莱姆法则推出这一结论。$d = 0$ 这一事实意味着对所有的 $j$，$|A_j|$ 一定包含一个全部元素为零的列，因而解值将为

$$\bar{x}_j = \frac{|A_j|}{|A|} = \frac{0}{|A|} = 0 \qquad (j = 1, 2\cdots, n)$$

令人费解的是，齐次方程组得到**非平凡解**的唯一方式是有 $|A| = 0$，即具有**奇异**矩阵 $A$！在此情况下我们有

111

$$\bar{x}_j = \frac{|A_j|}{|A|} = \frac{0}{0}$$

① “齐次”一词描述这样一种性质：当所有变量 $x_1, \cdots, x_n$ 被一常数乘时，方程组仍然成立。仅当常数项（不属于任意 $x_i$ 的项）全部为零时，才有这种可能。

这里表达式 0/0 并不等于 0,而是没有定义。因而克莱姆法则不能应用。这并不意味着我们不能得到解,而是意味着我们不能得到唯一解。

考察齐次方程组

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= 0 \\ a_{21}x_1 + a_{22}x_2 &= 0 \end{aligned} \tag{5.16}$$

很明显,$\bar{x}_1=\bar{x}_2=0$ 是一个解,但这个解为平凡解。现在假定系数矩阵 $A$ 为非奇异矩阵,所以 $|A|=0$。这意味着行向量$[a_{11}\ a_{12}]$为行向量$[a_{21}\ a_{22}]$的倍数,结果其中一个方程是多余的。假如从(5.16)中删除第二个方程,我们得到一个(第一个)方程,具有两个变量,解是 $\bar{x}_1=(-a_{12}/a_{11})\bar{x}_2$。此解为非平凡解,且若 $a_{11}\neq 0$,此解有定义,但实际上它代表了无穷多个解,因为对于每一个可能的 $\bar{x}_2$ 值,都存在一个对应的 $\bar{x}_1$ 值,使得 $\bar{x}_1$ 和 $\bar{x}_2$ 的解的关系成立。因此对于齐次方程组,不存在唯一的非平凡解。这个结论对于几个变量的情况也是成立的。

## 线性方程组解的结果

我们对线性方程组 $Ax=d$ 的几种变式的讨论表明,其解的结果可能存在四种类型。为更好地全面考察这些变式,我们将其列于表 5.1 中

在第一种情况下,方程组可得到惟一的非平凡解。仅在方程 112 组为非齐次方程组,其系数矩阵为非奇异矩阵的情况下,我们才可能得到这一结果。第二种可能的结果是唯一的平凡解。这一结果与具有非奇异矩阵 $A$ 的齐次方程组相联系。在第三种可能性中,

我们可能具有无数个解。这种偶然性仅当方程组中方程间存在相关性(即存在多余方程)时才会出现。平凡解是否包括在无数解中,则要视方程组是否为齐次方程组而定。最后,在方程组不相容的情况下,根本没有解。从模型构建者的角度看,最有益、最理想的结果当然是存在唯一的非平凡解 $\bar{x}\neq 0$。

**表 5.1　线性方程组 $Ax=d$ 解的结果**

| 行列式 $\lvert A\rvert$ ＼ 向量 $d$ | | $d\neq 0$ (非齐次方程组) | $d=0$ (齐次方程组) |
|---|---|---|---|
| $\lvert A\rvert\neq 0$ A 为非奇异矩阵 | | 存在唯一的非平凡解 $\bar{x}\neq 0$ | 存在唯一的平凡解 $\bar{x}=0$ |
| $\lvert A\rvert=0$ A 为奇异矩阵 | 方程相关 | 存在无数个解(不包括平凡解) | 存在无数个解(包括平凡解) |
| | 方程不相容 | 无解 | 不能应用 |

## 练习 5.5

1　运用克莱姆法则解下列方程组

(a) $3x_1-2x_2=11$
$2x_1+x_2=12$

(b) $-x_1+3x_2=-3$
$4x_1-x_2=12$

(c) $8x_1-7x_2=-6$
$x_1+x_2=3$

(d) $6x_1+9x_2=15$
$7x_1-3x_2=4$

2　求上题每一方程组系数矩阵的逆,并以公式 $\bar{x}=A^{-1}d$ 求方程组的解。

3　运用克莱姆法则解下列方程组

(a) $8x_1-x_2=15$
$x_2+5x_3=1$

(c) $4x+3y-2z=7$
$x+y=5$

$$
\begin{array}{ll}
2x_1 + 3x_3 = 4 & 3x + z = 4 \\
(b)\quad -x_1 + 3x_2 + 2x_3 = 24 & (d)\quad -x + y + z = a \\
x_1 + x_3 = 6 & x - y + z = b \\
5x_2 - x_3 = 8 & x + y - z = c
\end{array}
$$

4　证明:克莱姆法则还可以按下列步骤推导出来:以余子式 $|C_{1j}|$ 同时乘方程组 $Ax = d$ 中第一个方程的两边,再以余子式 $|C_{2j}|$ 同乘第二个方程的两边,等等。将所有新得到的方程相加,然后连续赋下标 $j$ 以数值 1,2,…,$n$,这样便得到如(5.14)所示的解值 $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n$。

---

## 5.6 克莱姆法则在市场模型和国民收入模型中的应用

第 3 章中讨论的简单均衡模型可通过克莱姆法则或矩阵求逆的方法很容易地解出。

### 市场模型

在消去数量变量以后,(3.12)所描述的两商品模型可以像(3.13′)那样写成两个线性方程组成的方程组:　113

$$
\begin{aligned}
c_1 P_1 + c_2 P_2 &= -c_0 \\
\gamma_1 P_1 + \gamma_2 P_2 &= -\gamma_0
\end{aligned}
$$

所需的三个行列式 $|A|$,$|A_1|$,$|A_2|$ 的值如下:

$$
|A| = \begin{vmatrix} c_1 & c_2 \\ \gamma_1 & \gamma_2 \end{vmatrix} = c_1\gamma_2 - c_2\gamma_1
$$

$$
|A_1| = \begin{vmatrix} -c_0 & c_2 \\ -\gamma_0 & \gamma_2 \end{vmatrix} = -c_0\gamma_2 + c_2\gamma_0
$$

$$|A_2| = \begin{vmatrix} c_1 & -c_0 \\ \gamma_1 & -\gamma_0 \end{vmatrix} = -c_1\gamma_0 + c_0\gamma_1$$

因此均衡价格必为

$$\bar{P}_1 = \frac{|A_1|}{|A|} = \frac{c_2\gamma_0 - c_0\gamma_2}{c_1\gamma_2 - c_2\gamma_1}$$

$$\bar{P}_2 = \frac{|A_2|}{|A|} = \frac{c_0\gamma_1 - c_1\gamma_0}{c_1\gamma_2 - c_2\gamma_1}$$

它们与(3.14)和(3.15)式中的结果完全一致。同以往一样,在需求或供给方程中通过设定 $P_1 = \bar{P}_1$, $P_2 = \bar{P}_2$,便可求得均衡数量。

## 国民收入模型

(3.23)式所描述的简单国民收入模型也可用克莱姆法则来解。如(3.23)所示,模型包括下列两个联立方程组:

$$Y = C + I_0 + G_0$$
$$C = a + bY \qquad (a > 0,\ 0 < b < 1)$$

可将其重排为

$$Y - C = I_0 + G_0$$
$$-bY + C = a$$

从而使内生变量 $Y$ 和 $C$ 仅出现在等号左边,而外生变量和常数参数位于等号的右边。系数矩阵的形式为 $\begin{bmatrix} 1 & -1 \\ -b & 1 \end{bmatrix}$,常数列向量(数据)为 $\begin{bmatrix} I_0 + G_0 \\ a \end{bmatrix}$。注意,和 $I_0 + G_0$ 被视作一个整体,即常数向量中的一个元素。



利用克莱姆法则可立即得到下列解

$$\bar{Y} = \frac{\begin{vmatrix} (I_0 + G_0) & -1 \\ a & 1 \end{vmatrix}}{\begin{vmatrix} 1 & -1 \\ -b & 1 \end{vmatrix}} = \frac{I_0 + G_0 + a}{1 - b}$$

$$\bar{C} = \frac{\begin{vmatrix} 1 & (I_0 + G_0) \\ -b & a \end{vmatrix}}{\begin{vmatrix} 1 & -1 \\ -b & 1 \end{vmatrix}} = \frac{a + b(I_0 + G_0)}{1 - b}$$

读者可以验证,刚得到的这个解与(3.24)和(3.25)式中的解是一致的。

现在我们试用系数矩阵求逆的方法来解此模型。因为系数矩阵为 $A = \begin{bmatrix} 1 & -1 \\ -b & 1 \end{bmatrix}$,其余子式矩阵 $\begin{bmatrix} 1 & b \\ 1 & 1 \end{bmatrix}$,因而我们有 $\operatorname{adj} A = \begin{bmatrix} 1 & 1 \\ b & 1 \end{bmatrix}$。由此可知逆矩阵为

$$A^{-1} = \frac{1}{|A|} \operatorname{adj} A = \frac{1}{1 - b} \begin{bmatrix} 1 & 1 \\ b & 1 \end{bmatrix}$$

我们知道,对方程组 $Ax = d$,解可以表示为 $\bar{x} = A^{-1} d$。将其应用于现在的模型,这意味着

$$\begin{bmatrix} \bar{Y} \\ \bar{C} \end{bmatrix} = \frac{1}{1 - b} \begin{bmatrix} 1 & 1 \\ b & 1 \end{bmatrix} \begin{bmatrix} I_0 + G_0 \\ a \end{bmatrix} = \frac{1}{1 - b} \begin{bmatrix} I_0 + G_0 + a \\ b(I_0 + G_0) + a \end{bmatrix}$$

很容易看到,这个解同前面得到的解也是相同的。

## 矩阵代数与变量消去法

这里举例说明的两个经济模型均仅包含两个方程,所以只需

计算二阶行列式。对于大的方程组,将会出现高阶行列式;而高阶行列式的计算并非轻松的任务。大逆阵的求逆更是绝非儿戏。事实上,从计算的角度看,克莱姆法则和矩阵求逆并不必然比连续变量消去法更有效。

有人可能会问,既然如此,为什么要使用矩阵方法呢?正如我们在前面的内容中所看到的,矩阵代数使我们可以用简洁的符号表示任意方程组,还给出了检验方程组存在唯一解的行列式判别准则。这些优点是其它方法不具备的。此外,需要提到的是,消元法不能给出任何解的分析表达式,而矩阵求逆和克莱姆法则却与其不同,它们能给出解的简洁表达式 $\bar{x} = A^{-1}d$ 和 $\bar{x}_j = |A_j| / |A|$。这种解的分析表达式的用途不仅在于它们是实际求解过程的归纳总结,而且也在于在需要时可对这些解进行进一步的数学运算。

115

在某些情况下,矩阵方法甚至具有计算上的优势。比如当同时解几个具有相同系数矩阵 $A$ 但常数项向量不同的方程组时,就有这一优点。在此情况下,考虑每个新方程组时,若用变量消去法每次均需重复这一计算过程。但用矩阵求逆法,仅需一次求出共同的逆矩阵 $A^{-1}$,然后用其左乘属于不同方程组的所有常数项向量,以分别得到不同方程组的解值。在我们下一节考察里昂惕夫投入—产出模型时,这种计算上的特殊优点具有极大的实用价值。

## 练习 5.6

1 用下列方法解练习 3.5—1 的国民收入模型:

($a$) 矩阵求逆法 ($b$) 克莱姆法则

(按顺序 $Y$、$C$、$T$ 排列变量)

2 用下列方法解练习 3.5—2 的国民收入模型:

($a$) 矩阵求逆法 ($b$) 克莱姆法则

(按顺序 $Y$、$C$、$G$ 排列变量。)

---

## 5.7 里昂惕夫投入—产出模型

里昂惕夫教授的"静态"投入—产出分析[①] 研究这一特殊问题:"在一个 $n$ 部门(产业)经济中的每一部门维持何种产出水平,才能充分满足对这种产品的总需求?"

投入—产出分析这一思想的合理性是很明显的。任何一个产业(比如钢铁工业)的产出,往往是其它许多产业的投入,或者是该产业自身的投入。因此,"正确"(即既不短缺也不滞存)的钢铁产出水平将取决于所有几个产业的投入需求。而其它许多产业的产出又是钢铁工业的投入,因而其它产业部门"正确"的产业水平又部分地取决于钢铁工业的投入需求。考虑到产业之间的相关性,几个产业"正确"的产出水平必定是与该经济所有投入需求相一致 116 的产出水平,从而不存在任何"瓶颈"现象。由此看来,投入—产出分析在制订生产计划过程中是极有价值的。比如,它对一国制订经济发展规划或国防规划,都有很大用处。

严格地讲,投入—产出分析不是我们在第三章中所讨论的一

---

① 瓦西里·里昂惕夫:《1919—1939 年美国经济结构》,第二版,牛津大学出版社,1951 年。

般均衡分析。尽管这里也强调不同产业之间的内在联系，但所设想的“正确”的产出水平是为了满足技术上的投入—产出关系，而不是为了满足市场均衡条件。然而，投入—产出分析所提出的问题也可归结为解联立方程组的问题，所以矩阵代数又是很有帮助的。

## 投入—产出模型的结构

因为一个投入—产出模型通常涵盖很多产业部门，其结构是非常复杂的。为使问题简化，我们采用下列假设：(1) 每个产业仅生产一种同质的产品（从广义上解释，两种或两种以上联合生产的产品，若它们彼此是按固定比例生产的，也是允许的）。(2) 每个产业用固定的投入比例或要素组合生产其产品，并且(3) 每一产业的生产面临固定的规模收益，从而所有投入增加 $k$ 倍，产出也将恰好增加 $k$ 倍。当然，这些假设是不现实的，但这一缺点也有可弥补之处：因为一个产业者生产两种不同的商品或使用两种不同的要素组合，则至少从概念上可以将该产业分成两个独立的部门。

由上述假设可知，为生产每一单位 $j$ 产品所需的投入的第 $i$ 种商品一定为一固定数量，我们以 $a_{ij}$ 表示。具体而言，生产每单位 $j$ 商品需 $a_{1j}$ 的第一种商品，$a_{2j}$ 的第二种商品，……，$a_{nj}$ 的第 $j$ 种商品。（$a_{ij}$ 的下标的顺序是容易记忆的：第一个下标是指投入，第二个下标是指产出，因而 $a_{ij}$ 表示生产每单位的 $j$ 产品，需要投入多少 $i$ 商品。）基于我们的目的，我们可以假设价格是给定的，并采用美元作为每种商品的计价单位。这样，$a_{32} = 0.32$ 表明生产 1 美元的第二种商品需要投入 35 美分的第三种商品。符号 $a_{ij}$ 可以称作投入系数。

对于 $n$ 部门经济，投入系数可排成表 5.2 那样的矩阵 $A =$

$[a_{ij}]$，其中每一列表示生产每单位特定产业的产品所需的投入。例如，第二列表示生产一单位(1 美元)的第 II 种商品，所需投入的第 I 种商品为 $a_{12}$单位，第 II 种商品为 $a_{22}$单位，等等。若每一产业均不需本产业的产品作为投入，则矩阵 $A$ 主对角线上所有元素均为零。 117

表 5.2　　投入系数矩阵

| 输入 | 输出 | | | | |
|---|---|---|---|---|---|
| | I | II | III | … | N |
| I | $a_{11}$ | $a_{12}$ | $a_{13}$ | … | $a_{1n}$ |
| II | $a_{21}$ | $a_{22}$ | $a_{23}$ | … | $a_{2n}$ |
| III | $a_{31}$ | $a_{32}$ | $a_{33}$ | … | $a_{3n}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | | ⋮ |
| N | $a_{n1}$ | $a_{n2}$ | $a_{n3}$ | … | $a_{nn}$ |

## 开放模型

除了 $n$ 个产业部门以外，如果模型中还包含一个“开放”的部门(比如居民户)，它外生地决定了对每个产业产品的**最终需求**(非投入需求)，而且它还供给并非由 $n$ 个产业生产的**基本投入**(比如劳务)，则此模型便为开放模型。

考虑到开放部门的存在，投入系数矩阵 $A$(或简称投入矩阵 $A$)每一列的元素和必定小于 1。每一列的和代表生产价值为 1 美元的某种商品所需的**部分**投入成本(不包括基本投入成本)。若此和大于或等于 1 美元，则该生产在经济上是不合算的。这一事实可用符号表示如下：

$$\sum_{i=1}^{n} a_{ij} < 1 \qquad (j = 1,2,\cdots,n)$$

这里是对 $i$ 求和，即对某一特定的 $j$ 列出现于不同行的元素求和。沿此思路进一步考虑，还可以这样推论；因为产出的价值(1 美元)一定被对所有生产要素的支付所吸收，某一列之和不足 1 的量(即 1 与该列和之差)必定代表对开放部门基本投入的支付。因此，生产 1 单位 $j$ 商品所需的基本投入值应为 $1-\sum_{i=1}^{n} a_{ij}$。

若产业 I 要生产恰好足以满足 $n$ 个产业的投入需求以及开放部门最终需求的产品，其产出水平 $x_1$ 必定满足下列方程：

$$x_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + d_1$$

或
$$(1-a_{11})x_1 - a_{12}x_2 - \cdots - a_{1n}x_n = d_1$$

118 其中 $d_1$ 表示对其产出的最终需求，$a_{1j}x_j$ 代表第 $j$ 产业的投入需求。[①] 注意，除了第一个系数 $(1-a_{11})$，后一方程其余的系数均是从表 5.2 的第一行直接移植下来的，不同的是在这里它们的前面均冠以负号。类似地，对应于产业 II 的方程具有与表 5.2 第二行相同的系数(也要加上负号)，但变量 $x_2$ 的系数为 $(1-a_{22})$ 而非 $-a_{22}$。对于全部 $n$ 个产业的集合，“正确”的产出水平可用下列 $n$ 个方程组来概括：

$$\begin{aligned}
(1-a_{11})x_1 - \quad a_{12}x_2 - \cdots - \quad a_{1n}x_n &= d_1 \\
-a_{21}x_1 + (1-a_{22})x_2 - \cdots - \quad a_{2n}x_n &= d_2 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
-a_{n1}x_1 - \quad a_{n2}x_2 - \cdots + (1-a_{nn})x_n &= d_n
\end{aligned} \tag{5.17}$$

① 不要把投入系数按行加起来，比如按 $a_{11}+a_{12}+\cdots+a_{1n}$ 方式加起来，这种加和没有经济意义。另一方面，乘积和 $a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n$ 确实具有经济意义，它表示作为所有 $n$ 个产业部门的投入，所需的 $x_1$ 的总量。

用矩阵符号，可以将其表示成：

$$\begin{bmatrix} (1-a_{11}) & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & (1-a_{22}) & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & (1-a_{nn}) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} \tag{5.17'}$$

若省略矩阵主对角线元素中所有的 1，则可将矩阵简写为 $-A=[-a_{ij}]$。另一方面，这个矩阵实际上是单位矩阵 $I_n$（主对角元素为 1，其余元素为 0）与矩阵 $-A$ 之和，因而(5.17′)也可以写成

$$(I-A)x=d \tag{5.17''}$$

其中 $x$ 和 $d$ 分别为变量向量和最终需求（常数项）向量。矩阵 $(I-A)$ 被称作**技术矩阵**，以 $T$ 表示。所以方程组也可以写成

$$Tx=d \tag{5.17'''}$$

只要 $T$ 为非奇异矩阵（没有事先的理由认定它不是非奇异矩阵），则可求其逆 $T^{-1}$，并由下列方程求得该方程组的唯一解：

$$\bar{x}=T^{-1}d=(I-A)^{-1}d \tag{5.18}$$

## 一个数字例子

为举例说明，假定一个经济体中只有三个产业，其投入系数 119
矩阵如下（这里使用小数）：

$$A=\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 0.2 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.2 \end{bmatrix} \tag{5.19}$$

注意,$A$ 中每列的和小于 1,它也应该小于 1。进而,若我们以 $a_{0j}$ 表示生产 1 美元的 $j$ 产品所需的基本投入的美元数量,用 1 减(5.19)每一列的和,我们有

$$a_{01}=0.3 \qquad a_{02}=0.3 \qquad 和 \qquad a_{03}=0.4 \tag{5.20}$$

运用上述矩阵 $A$,以 $Tx=(I-A)x=d$ 形式表示开放的投入—产出系统如下:

$$\begin{bmatrix} 0.8 & -0.3 & -0.2 \\ -0.4 & 0.9 & -0.2 \\ -0.1 & -0.3 & 0.8 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}=\begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \tag{5.21}$$

这里我们有意不给出最终需求 $d_1,d_2,d_3$ 的值,通过把向量 $d$ 保持成参数形式,其解便会以"公式"形式出现,给出各种具体的 $d$ 向量值,我们就可以得到各种相应的具体解值。

通过求出 3×3 技术矩阵 $T$ 的逆,可以近似地求出(5.21)的解(因小数的四舍五入)为:

$$\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{bmatrix}=T^{-1}d=\frac{1}{0.384}\begin{bmatrix} 0.66 & 0.30 & 0.24 \\ 0.34 & 0.62 & 0.24 \\ 0.21 & 0.27 & 0.60 \end{bmatrix}\begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$$

若具体的最终需求向量(比如,某一开发项目的最终产出目标)恰好为 $d=\begin{bmatrix} 10 \\ 5 \\ 6 \end{bmatrix}$,单位为十亿美元,则可得到下面具体的解值(单位同样为十亿美元):

$$\bar{x}_1=\frac{1}{0.384}[0.66(10)+0.30(5)+0.24(6)]=\frac{9.54}{0.384}=24.84$$

类似地

$$\bar{x}_2 = \frac{7.94}{0.384} = 20.68 \qquad 和 \qquad \bar{x}_3 = \frac{7.05}{0.384} = 18.36$$

现在产生一个重要问题。产出混合 $\bar{x}_1$，$\bar{x}_2$，$\bar{x}_3$ 的生产必定需要一定数量的基本投入。那么，一个经济所需要的基本投入数 120 量与其能够供给的数量是否一致呢？在(5.20)式基础上，所需的基本投入可计算如下：

$$\sum_{j=1}^{3} a_{0j}\bar{x}_j = 0.3(24.84) + 0.3(20.68)$$
$$+ 0.4(18.36) = \$21.00(10\text{ 亿})$$

因而当且仅当适宜的基本投入数量不少于 210 亿美元时，具体的最终需求向量 $d = \begin{bmatrix} 10 \\ 5 \\ 6 \end{bmatrix}$ 才是可行的。如果适宜的数量下降，则特定的产出目标当然不得不相应地向下调整。

上述分析的一个重要特征是，只要投入系数保持不变，逆矩阵 $T^{-1} = (I - A)^{-1}$ 就将保持不变。因此，即使我们需要考察成百上千个不同的最终需求向量(像一个连续的发展目标谱)。也仅需计算一个逆矩阵。这意味着与变量消去法相比较，可以节约大量的计算，特别是涉及大方程组时更是如此。注意，克莱姆法则并不具备这一优点。按照克莱姆法则，应根据公式 $\bar{x}_j = |T_j| / |T|$ 求解，但每次都要使用不同的最终需求向量 $d$，因而必须重新计算行列式 $|T_j|$，这比将一个已知的逆矩阵 $T^{-1}$ 被新的向量 $d$ 乘更耗费时间。

## 用近似方法求逆矩阵

对于大方程组，求逆矩阵是一个极其繁杂的工作。尽管计算

机可以帮助我们,但相对简单的计算方法仍是需要的。对所考察的投入产出模型,确实存在一种求出逆矩阵 $T^{-1}=(I-A)^{-1}$近似值的方法,且可以精确到任意的程度。因而可能完全回避了矩阵求逆的方法。

我们首先考察下列矩阵积($m$ 为正整数):

$$
\begin{aligned}
&(I-A)(I+A+A^2+\cdots+A^m)\\
&\quad =I(I+A+A^2+\cdots+A^m)-A(I+A+A^2+\cdots+A^m)\\
&\quad =(I+A+A^2+\cdots+A^m)-(A+A^2+\cdots+A^m+A^{m+1})\\
&\quad =I-A^{m+1}
\end{aligned}
$$

如果乘积的结果仅为单位矩阵 $I$,我们就可以取矩阵和$(I+A+A^2+\cdots+A^m)$作为$(I-A)$的逆矩阵。麻烦就出在 $-A^{m+1}$这项上。幸运的是,我们还有次优的选择。因为若矩阵 $A^{m+1}$能趋近于 $n\times n$ 零矩阵,则 $I-A^{m+1}$将趋近于 $I$,相应地,所说的矩阵和 121 $(I+A+A^2+\cdots+A^m)$将趋近于所希望的逆矩阵$(I-A)^{-1}$。因而,通过 $A^{m+1}$趋近于零矩阵,我们便能够得到源于矩阵 $I,A,A^2,\cdots,A^m$ 的加和的**近似逆矩阵**。

但我们能够使 $A^{m+1}$趋近于零矩阵吗?若能够,怎样才能使其趋近于零矩阵呢?如果矩阵 $A$ 中每列元素为非负,且其和小于 1(所考察的投入产出模型正是如此),正如(5.19)所说明的那样,那么,第一个问题的答案就是肯定的。在此情况下,只要使幂 $m$ 足够大,即使矩阵 $A$ 重复自乘的过程足够长,$A^{m+1}$就可以趋近于零矩阵。对于这一论断,我们将给出证明,但若我们现在假设这一结论是有效的,则计算近似逆矩阵的步骤便很清楚了:我们逐次计算矩阵 $A^2,A^3,\cdots$,直至出现矩阵 $A^{m-1}$,其所有元素按预先选定的标准均达到可以忽略的程度("趋近于零")。至此,我们便可以终

止自乘过程，并将所得到的所有矩阵相加，以形成近似逆矩阵$(I+A+A^2+\cdots+A^m)$。①

注意，当 $m$ 无限增加，矩阵 $A$ 使得$A^{m+1}$趋近于零矩阵时，近似矩阵$(I+A+A^2+\cdots+A^m)$也具有所有元素非负的性质。和中的前两项，$I$ 和 $A$，明显仅包含非负元素。所以 $A$ 的所有次幂均也将仅包含非负元素。因为 $A$ 的自乘仅包含 $A$ 自身元素的相乘和相加。由于最终需求向量也仅包含非负元素，因而由(5.18)明显可知，解出的产出水平也将非负。当然，这也正是我们所希望的结果。

现在我们来简单地证明这个结论：给定投入系数矩阵 $A=[a_{ij}]$，其每列和小于1，当 $m$ 无限增加时，矩阵 $A^{m+1}$将趋近于零矩阵。② 为此，我们需要矩阵 $A$ 的**范数**的概念，矩阵 $A$ 的范数定义为 $A$ 中最大的列的元素和，以 $N(A)$表示。例如在(5.19)的矩阵中，我们有 $N(A)=0.7$。它是第一列的和，它也恰好等于第二列的和。很明显，一个矩阵中没有元素能够超过范数的值。即：

$$a_{ij}\leqslant N(A) \qquad (\text{对所有的 } i \text{ 和 } j)$$

在投入产出方面，我们有 $N(A)<1$，且所有的 $a_{ij}<1$。实际上，矩 122 阵 $A$ 的元素为非负，我们必定有

① 以$(I+A+A^2+\cdots+A^m)$作为$(I-A)^{-1}$的近似类似于下列无穷级数与$(1+r+r^2+\cdots+r^n)$的近似关系

$$(1-r)^{-1}=\frac{1}{1-r}=1+r+r^2+\cdots \qquad (0<r<1)$$

因为通过选择适当数字 $n$，使级数中后面的项逐渐变小，从而使$(1-r)^{-1}$的上述近似值精确到我们所要求的任意程度。

② 更详尽的讨论，请参见弗里德里克·V.沃："通过幂级数求里昂惕夫矩阵的逆"《经济计量学》，1990年4月，第142—154页。

$$0 < N(A) < 1$$

关于矩阵的范数,存在这样一个定理:给定任意两个矩阵 $A$ 和 $B$,其积矩阵 $AB$ 的范数决不会超过 $N(A)$ 与 $N(B)$ 之积:

$$N(AB) \leqslant N(A)\ N(B) \tag{5.22}$$

在 $A = B$ 的特殊情况下,其中矩阵为方阵,这个结论意味着

$$N(A^2) \leqslant [N(A)]^2 \tag{5.23}$$

当 $B = A^2$,(5.22)和(5.23)两式意味着

$$N(A^3) \leqslant N(A)N(A^2) \leqslant N(A)[N(A)]^2 = [N(A)]^3$$

最后一个结果的一般形式为

$$N(A^m) \leqslant [N(A)]^m \tag{5.24}$$

正是基于上述关系,$0 < N(A) < 1$ 才显得极其重要,因为若 $N(A)$ 为一正分数,当 $m$ 无限增大时,$[N(A)]^m$ 必然趋近于零。根据(5.24)式,这意味着 $N(A^m)$ 也将趋近于零,因为 $N(A^m)$ 充其量仅等于$[N(A)]^m$。但若如此,当 $m$ 无限增加时,矩阵 $A^m$ 中的元素同样将趋近于零,因为 $A^m$ 中每个元素的值均不超过范数 $N(A^m)$。所以当条件 $0 < N(A) < 1$ 满足时,只要 $m$ 充分大,矩阵 $A^{m+1}$ 便可趋近于零矩阵。

## 封闭模型

若投入产出模型中的外生部门被纳入该系统,并成为其中的一个产业部门,则该开放模型便变成**封闭模型**。在封闭模型中,不再出现最终需求和基本投入,其位置由新构想的产业部门的投入需求和产出所填补。所有物品均具有**中间产品**的性质,仅是为了满足 $(n+1)$ 个产业的投入需求而生产的。

粗略看来,开放部门转化成额外的产业部门似乎不会对分析

产生明显的变化。但实际上,因为新的产业被假定与其它产业一样具有固定的投入比例,过去用作基本投入的供给,现在必须与过去称作的最终需求保持固定的比例。更具体地说,比如,这意味着居民户将按照与其供给的劳务固定的比例消费每一商品。这当然使分析框架发生了明显的变化。

从数学角度看,最终需求的消失意味着我们将有一个齐次方程组。假设一经济中仅有四个产业(包括新的产业,以下标 0 标示),按(5.17′)类推,"正确的"产出水平应当满足下列方程组

123

$$\begin{bmatrix} (1-a_{00}) & -a_{01} & -a_{02} & -a_{03} \\ -a_{10} & (1-a_{11}) & -a_{12} & -a_{13} \\ -a_{20} & -a_{21} & (1-a_{22}) & -a_{23} \\ -a_{30} & -a_{31} & -a_{32} & (1-a_{33}) \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

因为该方程组是齐次的,当且仅当 4×4 技术矩阵$(I-A)$具有零行列式时,才会有非平凡解。后一条件实际上总是满足的:在一个封闭模型中,不再有基本投入,现在投入系数矩阵每一列的和必定等于 1(而非小于 1),即 $a_{0j}+a_{1j}+a_{2j}+a_{3j}=1$,或者

$$a_{0j}=1-a_{1j}-a_{2j}-a_{3j}$$

这意味着矩阵$(I-A)$的每一列的第一个元素总是等于其它三个元素之和的负值,结果,矩阵的四行线性相关,我们必然有 $|I-A|=0$。这确保了方程组确实具有非平凡解。事实上,如表 5.1 所指明的那样,它具有无穷多组解。这意味着在以齐次方程组表示的封闭模型中,不存在唯一"正确"的产出混合。我们可以确定产出水平 $\bar{x}_1, \cdots, \bar{x}_4$ 之间的比例,但不能固定产出的绝对水平,除非我们对模型施加额外的限制。

## 练习 5.7

---

1　基于(5.21)模型，若最终需求 $d_1=30$，$d_2=15$，$d_3=10$(单位均为 10 亿美元)，三个产业的产出水平解为多少？(保留两位小数。)

2　运用(5.20)式的信息，计算上题产出水平解所需的基本投入的全部数量。

3　在两产业部门经济中，产业 I 生产 1 美元的第 I 种商品需 10 美分的自身产品作投入，60 美分的第 II 产业的商品作投入。产业 II 生产 1 美元的 II 商品不需自身产业的产品作投入，仅需投入 50 美分的 I 商品。开放部门需要 10000 亿美元的商品 I，20000 亿美元的商品 II。

124

($a$) 写出投入矩阵，技术矩阵，该经济的投入产出矩阵方程。

($b$) 用克莱姆法则求产出水平解。

4　给定投入矩阵和最终需求向量

$$A=\begin{bmatrix}0.05 & 0.25 & 0.34\\ 0.33 & 0.10 & 0.12\\ 0.19 & 0.38 & 0\end{bmatrix}\quad d=\begin{bmatrix}1800\\ 200\\ 900\end{bmatrix}$$

($a$) 解释元素 0.33、0、200 的经济含义。

($b$) 充分解释第三列和的经济含义。

($c$) 充分解释第三行和的经济含义。

($d$) 写出该模型的具体投入产出矩阵方程。

5　用克莱姆法则解上题三个产业的产出水平解。

---

# 5.8　静态分析的局限性

在市场和国民收入静态均衡的讨论中，我们主要关心的是求模型中内生变量的均衡值。在这种分析中我们忽略了的一个基本

问题是:导致均衡状态的各变量调整和再调整的实际过程(若这种状态可以达到)。我们只关心在何处达到均衡状态,但对何时达到均衡状态以及达到均衡状态的过程中出现何种问题并不关心。

因此这种静态均衡分析有两个重要问题没有顾及。一个问题是,因为调整过程需很长时间方能完成,所以如果模型中的外生变量在此期间发生了某些变化,在特定静态分析框架中决定的静态均衡将在它最终达到之前,就失去了其实际意义。这就是均衡状态的移动问题。另一个问题是:即使调整的过程不受干扰,静态分析所构想的均衡状态也可能难以达到。这可能是由于出现了所谓的"非静态均衡"的情况,其特征是调整过程驱使变量逐渐偏离,而不是逐渐趋近于均衡状态。因此,忽略调整过程,也就是忽略了均衡的可实现性问题。

对由外生变化所致的均衡状态的转移的分析,属于**比较静态分析**的范畴;均衡的可实现性和稳定性问题则属于**动态分析**的范畴。很明显,比较静态分析和动态分析都将弥补静态分析的不足,所以探索这些领域也是我们极为迫切的任务。动态分析的研究将在本书第五篇进行,下面我们将注意力转移到比较静态分析的讨论上来。

# 第 三 篇

# 比较静态分析

# 第 6 章　比较静态学与导数的概念

127

本章和后面两章将专门研究比较静态分析的方法。

## 6.1　比较静态学的性质

比较静态学，顾名思义，涉及两种与不同参数值和内生变量相联系的不同均衡状态的比较。基于比较之目的，我们总是假设给定初始均衡状态作为出发点。例如，在孤立市场模型中，这样的初始均衡由确定的价格 $\bar{P}$ 和相应的数量 $\bar{Q}$ 代表。类似地，在(3.23)式的简单国民收入模型中，初始均衡由确定的 $\bar{Y}$ 和相应的 $\bar{C}$ 所规定。现在，若我们令模型发生非均衡变化(按某些参数或外生变量发生变化的形式进行)，则初始均衡必然会被打破。因而，各内生变量必然经历某些调整。若假设，与新的参数和变量值相关的新的均衡有定义且可达到，则比较静态分析面临的问题是：新均衡如何与原均衡相比较。

需指出的是，在比较静态分析中我们再一次忽略了变量的调整过程。我们仅比较初始(变化前的)均衡状态与最终(变化后的)均衡状态。同样，我们也回避了均衡的稳定性问题，因为我们像对待原均衡一样，假设新均衡是可实现的。

128

比较静态分析可以是定性分析，也可以是定量分析。比如，若

我们仅对投资 $I_0$ 的增加是增加还是减少均衡收入 $\bar{Y}$ 感兴趣，则这种分析是定性的，因为我们仅关心变化的方向。但若我们关注的是：由给定 $I_0$ 的变化，会导致 $\bar{Y}$ 变化的大小（即投资乘数），这种分析显然是定量的。而且通过获得定量的答案，由代数符号，我们可以自动地得知变化的方向。因而，定量分析总是包含定性分析。

很明显，我们关注的问题之一实质上是求变化率的问题：内生变量均衡值对特定参数或外生变量变化的比率。为此，导数这个数学概念在比较静态分析中具有极为重要的意义。因为导数这个在数学的一个分支——微分学中最基本的概念，直接与变化率这一概念有关。而且，我们后面将发现，导数这个概念对最优化问题也是极其重要的。

## 6.2 变化率与导数

尽管我们目前的讨论仅与模型中变量均衡值的变化率有关，但我们仍可以更一般的方式展开我们的讨论：考察任意变量 $y$ 对另一变量 $x$ 的变化率，其中这两个变量以下列函数联系起来

$$y = f(x)$$

将其应用于比较静态分析，变量 $y$ 将代表内生变量的均衡值，$x$ 代表某个参数。注意，在开始阶段，我们仅限于讨论模型中仅有一个参数或外生变量的情况。但一旦我们能够熟练处理这种简单情况，向更多参数的情况拓展就比较容易了。

### 差　商

因为“变化”这个概念现在极为重要，需要用一个特别的符号

来表示它。当变量 $x$ 由 $x_0$ 变化到 $x_1$ 时，这个变化由差 $x_1 - x_0$ 度量。因此，我们用符号 Δ（大写希腊字母，表示“差”）表示变化，129 写成 $\Delta x = x_1 - x_0$。我们还需要一种表示在不同 $x$ 值时函数 $f(x)$ 值的方式。标准的作法是使用 $f(x_i)$ 表示当 $x = x_i$ 时 $f(x)$ 的值。因此，对于函数 $f(x) = 5 + x^2$，我们有 $f(0) = 5 + 0^2 = 5$。同样，$f(2) = 5 + 2^2 = 9$，等等。

当 $x$ 由初始值 $x_0$ 变为新值 $(x_0 + \Delta x)$ 时，函数 $y = f(x)$ 的值由 $f(x_0)$ 变为 $f(x_0 + \Delta x)$。每单位 $x$ 的变化所引致的 $y$ 的变化可用**差商**来表示：

$$\frac{\Delta y}{\Delta x} = \frac{f(x_0 + \Delta_x) - f(x_0)}{\Delta x} \tag{6.1}$$

此差商度量 $y$ 的平均变化率。如果我们知道 $x$ 的初始值或 $x_0$，以及 $x$ 变化的大小或者 $\Delta x$，就可以对差商进行计算。即 $\Delta y / \Delta x$ 是 $x_0$ 和 $\Delta x$ 的函数。

**例 1** 给定 $y = f(x) = 3x^2 - 4$，我们可以写成：

$$f(x_0) = 3(x_0)^2 - 4 \qquad f(x_0 + \Delta x) = 3(x_0 + \Delta x)^2 - 4$$

因而，差商为

$$\frac{\Delta y}{\Delta x} = \frac{3(x_0 + \Delta x)^2 - 4 - (3x_0^2 - 4)}{\Delta x}$$

$$= \frac{6x_0\Delta x + 3(\Delta x)^2}{\Delta x} = 6x_0 + 3\Delta x \tag{6.2}$$

若已知 $x_0$ 和 $\Delta x$，则可计算差商。令 $x_0 = 3$，$\Delta x = 4$，则 $y$ 的平均变化率为 $6(3) + 3(4) = 30$。这意味着当 $x$ 由 3 变至 7 时，$x$ 每变化 1 个单位，$y$ 平均变化 30 个单位。

## 导　数

通常我们关心的是当 $\Delta x$ 很小时，$y$ 的变化率。在此情况下，有可能舍弃差商中涉及表达式 $\Delta x$ 的所有项，从而得到 $\Delta y/\Delta x$ 的近似值。例如在(6.2)中，若 $\Delta x$ 很小，我们可以仅取方程右侧的项 $6x_0$ 作为 $\Delta y/\Delta x$ 的近似。当然，$\Delta x$ 的值越小，近似值越趋近于 $\Delta y/\Delta x$ 的真实值。

当 $\Delta x$ 趋近于零(即 $\Delta x$ 越来越接近于零但从来不会成为零)，$(6x_0+3\Delta x)$将趋近于 $6x_0$，同样 $\Delta y/\Delta x$ 也将趋近于 $6x_0$。这一事实用符号可以表示成当 $\Delta x\to 0$ 时，$\Delta y/\Delta x\to 6x_0$，或者用下列方程表示

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}(6x_0+3\Delta x)=6x_0 \tag{6.3}$$

其中符号 $\lim\limits_{\Delta x\to 0}$ 读作："当 $\Delta x$ 趋近于零时，……的极限。"若当 $\Delta x\to 0$ 时，差商 $\Delta y/\Delta x$ 的极限存在，则该极限就是函数 $y=f(x)$的导数。

关于导数，有以下几点需要注意。首先，导数是一个函数，实 130 际上使用导数一词原义是指导函数。原来的函数 $y=f(x)$是原函数，而导数则是由原函数推导出的另一个函数。读者由(6.3)可知，差商是 $x_0$ 和 $\Delta x$ 的函数，但导数只是 $x_0$ 的函数。这是因为 $\Delta x$ 已被强制趋于零，因而它不能再被看作函数中的一个变量。我们还需补充说明，我们一直使用加下标的符号 $x_0$，仅是为了强调 $x$ 的变化必须从某个特定的 $x$ 值开始。既然我们已了解这一点，就可以去掉下标，并可以说导数像原函数一样，本身也是自变量 $x$ 的函数。即对于每个 $x$ 值，都有一个唯一的导函数值与之相

对应。

其次，因导数仅为差商之极限，而差商则度量 $y$ 的变化率，则导数必然也度量某种变化率。但由于导数中自变量 $x$ 的变化为无穷小($\Delta x \to 0$)，导数所度量的变化率具有瞬时变化率的性质。

第三，关于导数的记号问题，导函数通常用两种方法表示。给定原函数 $y=f(x)$，表示导数的一种方法是使用符号 $f'(x)$，或仅记为 $f'$。这种记法要归功于数学家拉格朗日。另一种常见记法是 $dy/dx$，由数学家莱布尼兹发明。[实际上还存在第三种记法：$Dy$ 或 $Df(x)$，但我们在后面的讨论中不使用这种记号。]记号 $f'(x)$与原函数的记号 $f(x)$相似，具有能表示导数本身是 $x$ 函数的优点。将导数表示成 $f'(x)$而非 $\phi(x)$，是为了强调函数 $f'$是由原函数 $f$ 推导而来的。另一种记号 $dy/dx$ 则是为了强调导数值能够度量变化率。字母"$d$"对应于希腊字母"$\Delta$"，$dy/dx$ 与 $\Delta y/\Delta x$ 的主要区别在于前者是后者当 $\Delta x \to 0$ 时的极限。在后面的讨论中，我们将依哪种符号更符合具体的情况而选择使用这两种符号。

运用这两种记号，我们可以将给定函数 $y=f(x)$的导数定义如下：

$$\frac{dy}{dx} \equiv f'(x) \equiv \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

**例 2** 我们已在(6.2)式中求出函数 $y=3x^2-4$ 的差商，差商的极限在(6.3)式中给出。根据(6.3)式，我们现在写出(以 $x$ 代替 $x_0$)：

$$\frac{dy}{dx}=6x \qquad \text{或} \qquad f'(x)=6x$$

注意，不同的 $x$ 值将对应不同的导数值。例如，当 $x=3$ 时，我们有 $f'(x)=6(3)=18$；但当 $x=4$ 时，我们求得 $f'(4)=6(4)=24$。[131]

**练习 6.2**

---

1 已知函数 $y=4x^2+9$：

($a$) 求作为 $x$ 和 $\Delta x$ 函数的差商。(使用 $x$ 代替 $x_0$)。

($b$) 求导数 $dy/dx$。

($c$) 求 $f'(3)$ 和 $f'(4)$。

2 已知函数 $y=5x^2-4x$：

($a$) 求作为 $x$ 和 $\Delta x$ 的函数的差商。

($b$) 求导数 $dy/dx$。

($c$) 求 $f'(2)$ 和 $f'(3)$。

3 已知函数 $y=5x-2$：

($a$) 求差商 $\Delta y/\Delta x$。它是何种类型的函数？

($b$) 因 $\Delta x$ 在上面的函数 $\Delta y/\Delta x$ 中并未出现，$\Delta x$ 的值变大或变小，对 $\Delta y/\Delta x$ 的值有影响吗？当 $\Delta x \to 0$ 时，差商的极限为多少？

---

## 6.3 导数与曲线的斜率

基础经济学告诉我们：给定总成本函数 $C=f(Q)$，其中 $C$ 表示总成本，$Q$ 表示产出，边际成本($MC$)定义为单位产出增加导致的总成本变化，即 $MC=\Delta C/\Delta Q$。当然 $\Delta Q$ 极小，在产品仅取离散单位(仅取整数)时，最小的变化就是一单位。但在产品数量为连续变量的情况下，$\Delta Q$ 是指无穷小的变化。众所周知，在后一种

情况下，边际成本可通过总成本曲线的斜率来度量。而总成本曲线的斜率只不过是当 $\Delta Q$ 趋于零时，比率 $\Delta C/\Delta Q$ 的极值。因而，曲线斜率的概念仅是导数概念的几何表示。二者均与经济学中广泛使用“边际”这一术语有关。

在图 6.1 中，我们绘出了原函数 $C=f(Q)$ 的图形，即总成本曲线 $C$。假设 $Q_0$ 为初始产出水平，并由此度量产出的增长，则在成本曲线上相应的点为 $A$。若产出增至 $Q_0+\Delta Q=Q_2$，则总成本将由 $C_0$ 增至 $C_0+\Delta C=C_2$，这样，$\Delta C/\Delta Q=(C_2-C_0)/(Q_2-$ 132 $Q_0)$。从几何上看，这是两条线段的比率，即 $EB/AE$，或者直线 $AB$ 的斜率。这个特定的比率度量平均变化率——即图中特定的

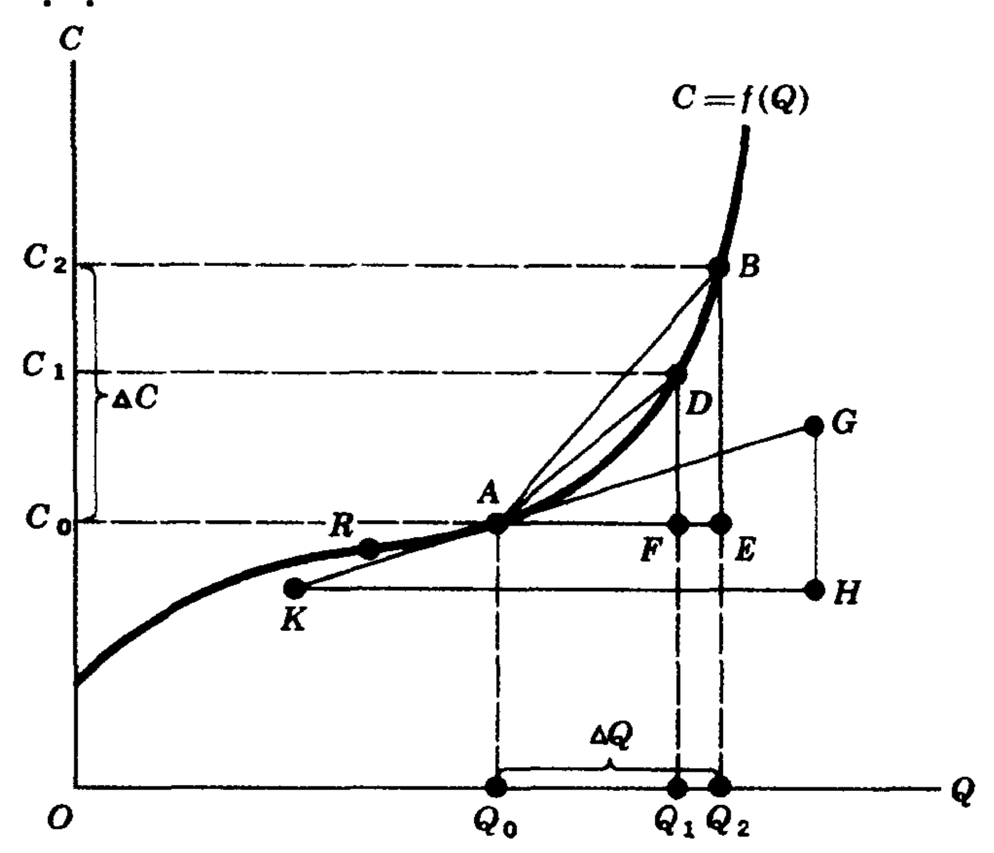



图 6.1

$\Delta Q$ 的平均边际成本——表示差商。所以,它是初始值 $Q_0$ 和变化量 $\Delta Q$ 的函数。

当我们改变 $\Delta Q$ 的大小,会出现何种情况?若我们仅考察较小的产出增量(如仅由 $Q_0$ 增至 $Q_1$),则平均边际成本将由直线 $AD$ 的斜率来度量。进而,随着我们不断减少产出增量,直线将越来越趋于平坦,直至极限即 $\Delta Q \to 0$ 时,我们得到直线 $KG$,即成本曲线在点 $A$ 的切线。$KG$ 的斜率 $HG/KH$,度量总成本曲线在点 $A$ 的斜率,并表示当 $\Delta Q \to 0$ 时 $\Delta C/\Delta Q$ 的极值(初始产出水平为 $Q = Q_0$)。因此,用导数的概念来表示,曲线 $C = f(Q)$ 在点 $A$ 的斜率对应于特定的导数值 $f'(Q_0)$。

若初始产出水平由 $Q_0$ 变到 $Q_2$,会出现何种情况?在此情况下,曲线上的点 $B$ 将取代点 $A$ 成为相关的点,曲线上新的点 $B$ 的斜率给出导数值 $f'(Q_2)$。对于其它初始产出水平,也可以得到类似的结果。一般而言,导数 $f'(Q)$ 作为 $Q$ 的函数,将随 $Q$ 的变化而变化。

## 6.4 极限的概念

导数 $dy/dx$ 被定义为当 $\Delta x \to 0$ 时,差商 $\Delta y/\Delta x$ 的极限。若我们采用简写符号 $q \equiv \Delta y/\Delta x$($q$ 表示差商)和 $v = \Delta x$($v$ 表示变差),则我们有:

133

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{v \to 0} q$$

鉴于导数概念极大地依赖于极限概念这一事实,我们迫切需要掌握极限这一概念。

### 左极限和右极限

极限这一概念与下述问题有关:"当一个变量,比如 $v$ 趋近于某一特定值,比如零时,另一个变量,比如 $q$,将趋近于何值?"要使这一问题有意义,$q$ 当然得是 $v$ 的函数,即 $q=g(v)$。我们迫切关注的是求出当 $v \to 0$ 时,$q$ 的极限。但我们前面应探讨更一般的情况,即当 $v \to N$,$N$ 为有限实数的情况。若如此,则$\lim\limits_{v \to 0} q$ 仅是$\lim\limits_{v \to N} q$,当 $N=0$ 时的一个特例。在讨论过程中,我们实际上也考察当 $v \to +\infty$(正无穷)或当 $v \to -\infty$(负无穷)时 $q$ 的极限。

当我们说 $v \to N$ 时,变量 $v$ 可以从大于$N$,也可以从小于 $N$ 的方向趋近于$N$。若当 $v$ 从左侧即小于$N$ 的方向趋于$N$,$q$ 趋近于一有限数$L$,则我们称 $L$ 为$q$ 的左极限;若当 $v$ 从右侧即大于 $N$ 的方向趋近于$N$ 时有极值$L$,我们则称 $L$ 为$q$ 的右极限。左极限和右极限可能相等,也可能不相等。

$q$ 的左极限以符号 $\lim\limits_{v \to N^-} q$(负号表示从小于 $N$ 的值趋近),右极限则写成 $\lim\limits_{v \to N^+} q$。当且仅当两极限具有共同的有限值,比如 $L$ 时,我们便认为 $q$ 的极限存在,并写成$\lim\limits_{v \to N} q=L$。注意,$L$ 必须为有限数值。若出现这种情况:$\lim\limits_{v \to N}=\infty$(或 $-\infty$),我们则认为 $q$ 无极限,因为$\lim\limits_{v \to N} q=\infty$意味着当 $v \to N$ 时,$q \to \infty$,如果当 $v$ 趋近于 $N$ 时,$q$ 的值永远增加,这同 $q$ 有极限的说法相矛盾。但作为表述当 $v \to N$ 时 $q \to \infty$这一事实的方便方法,人们实际上写成$\lim\limits_{v \to N} q=\infty$,并称 $q$ 具有"无穷大的极限"。

在某些情况下,只需考虑单侧极限。例如,在提到当 $v \to +\infty$

时 $q$ 的极限,仅有 $q$ 的左极限是可能存在的,因为 $v$ 仅从左侧才能趋于 $+\infty$。类似地,当 $v \to -\infty$ 时,仅有右极限才可能存在。在这些情况下,$q$ 的极限是否存在将取决于当 $v \to +\infty$ 或当 $v \to -\infty$ 时,$q$ 是否趋近于一个有限值。

认识到符号 $\infty$(无穷大)不是一个数,因而不能对其进行一般的代数运算是重要的。我们不能有 $3+\infty$ 或 $1/\infty$,我们不能写成 134 $q=\infty$,它不同于 $q \to \infty$。但把 $q$ 的极限表示成"="(相对于→)$\infty$ 是可以的,因为这仅表示 $q \to \infty$。

## 图例说明

在图 6.2 中,我们说明有关函数 $q=g(v)$ 的极限的几种可能的情形。

图 6.2$a$ 中的是一条平滑曲线。当变量 $v$ 从垂线两边中的一边趋近于 $N$ 值时,变量 $q$ 趋近于 $L$ 值。在此情况下,左极限与右极限的极限是一致的,因此,我们可以写成 $\lim\limits_{v \to N} q = L$。

图 6.2$b$ 中的是一条非平滑曲线。在 $N$ 点的正上方有一个尖转折点,但当 $v$ 从左右两边趋近于 $N$ 时,$q$ 同样趋近于同样的值 $L$。$q$ 的极限同样存在,并等于 $L$。

图 6.2$c$ 表示的是一条所谓的"阶段函数"。[①]在此情况下,当

① 阶段函数这个名称很容易通过曲线的形状来解释。但也可以通过代数方法来进行解释。图 6.2$c$ 所描述的情况可以用下述方程表示

$$q = \begin{cases} L_1 (0 \leqslant v < N) \\ L_2 (N \leqslant v) \end{cases}$$

注意,在上述定义域的每个子集中,函数表现为不同的常值函数,它们构成了"阶梯"形。

在经济学中,阶段函数可用于表示对不同的购买量所支付的不同价格(图 6.2$c$ 的曲线刻画出数量折扣的概念),也可以用于表示适用于不同收入等级的各种税率。

135

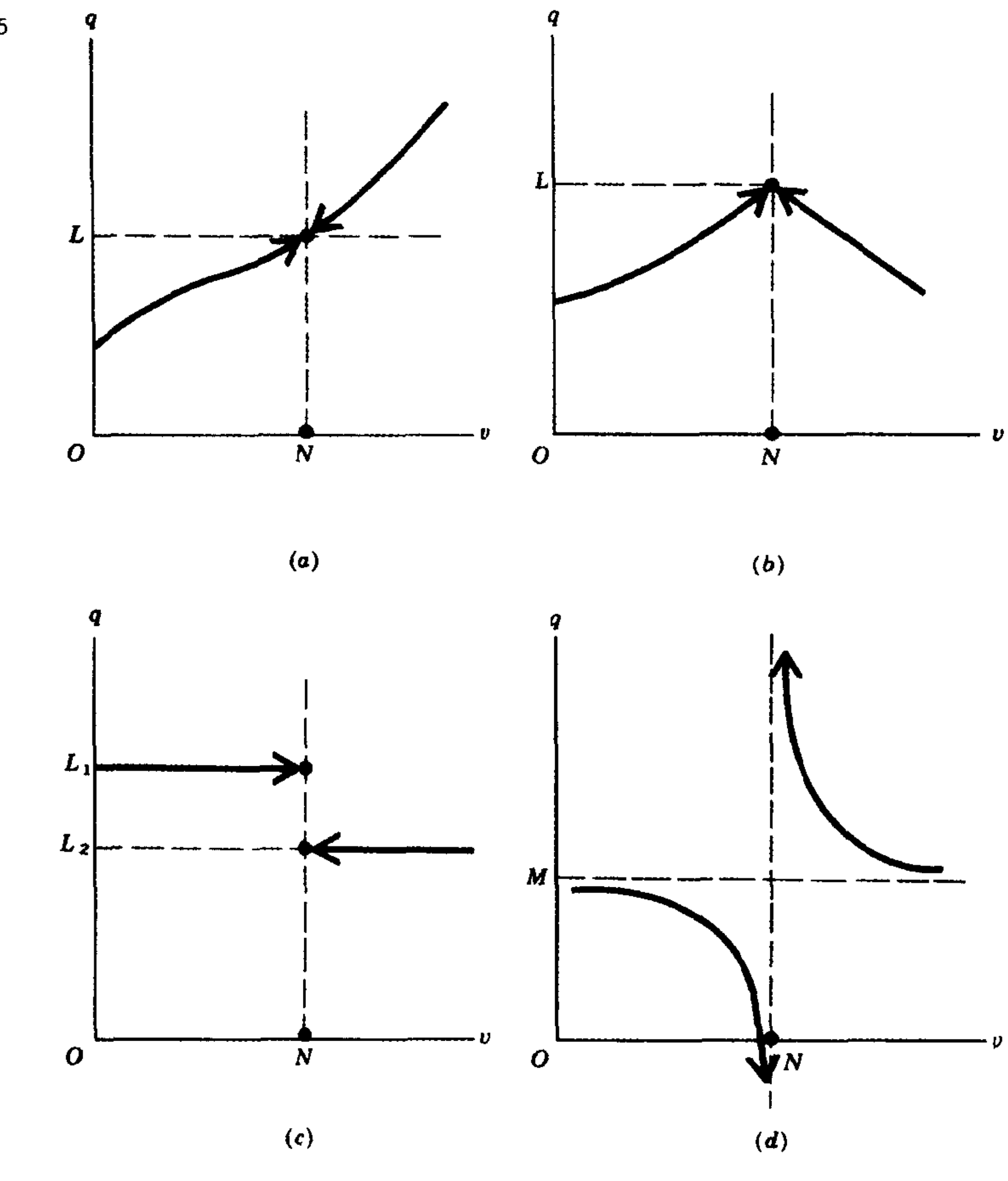



图 6.2

$v$ 趋近于 $N$ 时，$q$ 的左极限为 $L_1$，右极限为 $L_2$，二者数值不同。因此，当 $v \to N$ 时，$q$ 的极限不存在。

最后，在图 6.2$d$ 中，当 $v \to N$ 时，$q$ 的左极限为 $-\infty$，而右极限为 $+\infty$，这是因为双曲线的两部分与纵垂虚线即渐近线无限接

近时，它们将分别无限下降和上升。同样，$\lim\limits_{v \to N} q$不存在。另一方面，若我们考察图 6.2$d$ 中另一种类型的极限，即 $\lim\limits_{v \to \infty} q$，则与此相关的极限为左极限，且我们知道极限存在：$\lim\limits_{v \to +\infty} q = M$。类似地，读者还可以验证 $\lim\limits_{v \to -\infty} q = M$。

我们还可以用左、右极限的概念来讨论图 6.1 中的边际成本。在那里，变量 $q$ 和 $v$ 分别指差商 $\Delta C/\Delta Q$ 和 $\Delta Q$ 的大小，所有的变化均从曲线上的 $A$ 点开始度量。换句话说，$q$ 是指诸如 $AB$、$AD$、$KG$ 这类直线的斜率，而 $v$ 则指诸如线段 $Q_0Q_2(=AE)$、$Q_0Q_1(=AF)$等的长度。我们已经看到，当 $v$ 由正值趋于零时，$q$ 的值将趋近于$KG$ 的斜率。类似地，我们能够确定，如果 $\Delta Q$ 从负值趋于零时(即当产出越来越下降)，像 $RA$ 的斜率所度量的差商 $\Delta C/\Delta Q$，也将趋近于 $KG$ 的斜率。实际上，这非常类似于图 6.2$a$ 所描述的情况。图 6.1 中 $KG$ 的斜率(图 6.2 中 $L$ 的对应物)实际上是当 $v$ 趋于零时差商$q$ 的极限，因而它给出在产出水平 $Q=Q_0$ 时的边际成本。

### 极限的计算

136

现在我们举例说明给定函数 $q=g(v)$的代数运算。

**例 1** 已知 $q=2+v^2$，求$\lim\limits_{v \to 0} q$。为取左极限，我们以负值序列 $-1$、$-\frac{1}{10}$、$-\frac{1}{100}$……(依此顺序)替代 $v$，并求$(2+v^2)$，我们将看到，它将稳定地递减，并趋于2(因 $v^2$ 将逐渐趋于零)。其次，为求右极限，我们以正值序列 1、$\frac{1}{10}$、$\frac{1}{100}$、……(依此顺序)替代 $v$，并得到同前面一样的极限。由于左、右极限是一致的，我们认定 $q$ 的

极限存在，可写成$\lim_{v\to 0} q=2$。

始终存在这样一种诱惑，即将所得到的答案看作令方程 $q=2+v^2$中的 $v=0$ 的结果，但我们必须拒绝这种诱惑。在计算$\lim_{v\to N}$时，我们仅能令 $v$ 趋于$N$，但作为一个规则，我们一般不能令 $v=N$。实际上，即使 $N$ 不在函数$q=g(v)$的定义中，当 $v\to N$ 时，我们仍可以有充分理由提到 $q$ 的极限。在后一种情况下，若我们试图令 $v=N$，$q$ 显然无定义。

**例 2** 已知 $q=(1-v^2)/(1-v)$，求$\lim_{v\to 1} q$。这里，$N=1$ 不在函数的定义域中，我们不能令 $v=1$，因为这将使分母为零。而且，甚至令 $v\to 1$ 的极限计算步骤，像在例 1 中所运用的，也会导致一定困难，因为当 $v\to 1$ 时，分母$(1-v)$将趋于零，所以我们仍然无法进行除法运算。

克服这一困难的一种方法是对比率$(1-v^2)/(1-v)$进行变换，使分母中不再包含 $v$。因为 $v\to 1$ 意味着 $v\neq 1$，所以$(1-v)$不等于零，因而$(1-v^2)$被$(1-v)$除是可以的，并可写成[①]

$$q=\frac{1-v^2}{1-v}=1+v \qquad (v\neq 1)$$

---

① 像数的情形一样，可按下式进行除法：

$$\begin{array}{r|l} & 1+v \\ \hline 1-v & 1 \quad -v^2 \\ & \underline{1-v} \\ & v-v^2 \\ & \underline{\underline{v-v^2}} \end{array}$$

我们还可以进行如下因式分解：

$$\frac{1-v^2}{1-v}=\frac{(1+v)(1-v)}{1-v}=1+v \qquad (v\neq 1)$$

在 $q$ 的这个新表达式中,分母中不再包含 $v$。由于当 $v \to 1$ 时 137 $(1+v) \to 2$(左、右极限一致),所以我们可以得出这一结论:$\lim_{v \to 1} q = 2$。

**例 3** 已知 $q=(2v+5)/(v+1)$,求 $\lim_{v \to \infty} q$。这里,变量 $v$ 再次在分子和分母中同时出现。若我们令分子、分母中的 $v \to \infty$,则结果是两个无穷大数的比率,没有明确意义。为了克服这一困难,我们这次尝试将此给定比率进行变换,使分子中不再包含变量 $v$[①],这样,此给定比率相除,便可以求出极限。因 $(2v+5)$ 不能被 $(v+1)$ 整除,所以如下结果中含有一个余项:

$$q = \frac{2v+5}{v+1} = 2 + \frac{3}{v+1}$$

但无论如何,这个 $q$ 的新表达式分子中不再含有 $v$,注意,当 $v \to \infty$ 时,余项 $3/(v+1) \to 0$,所以我们可得出结论:$\lim_{v \to \infty} q = 2$。

极限的计算还有几个有用的定理,我们将在 6.6 节中对其进行讨论。

### 极限概念的正规考察

上面的讨论已介绍了极限概念的一般思想。现在我们给出更精确的定义。因为这一定义要用到直线上点(特别是作为实数线上的一个特定数的点)的**邻域**的概念,所以我们首先对其进行解释。

对于给定数 $L$,总会找到一个数 $(L-a_1)<L$ 和另一个数

---

① 注意,这不同于 $v \to 0$ 的情况。当 $v \to 0$ 时,我们将 $v$ 从分母中提出以免被零除。当 $v \to \infty$ 时,最好将 $v$ 从分子中提出。因为当 $v \to \infty$ 时,表达式分子中含有 $v$ 将变得无穷大,但若分母中含有 $v$,则将趋于零,从而平静地消失,使我们方便地求得极限。

$(L+a_2)>L$，其中 $a_1$ 和 $a_2$ 为任意正数，所有位于 $(L-a_1)$ 和 $(L+a_2)$ 之间的数的集合称作这两个数之间的区间，若数 $(L-a_1)$ 和 $(L+a_2)$ 包括在集合内，则此集合称作闭区间；若它们不包含在这集合中，此数集称作开区间。$(L-a_1)$ 和 $(L+a_2)$ 之间的闭区间用方括号表示

$$[L-a_1, L+a_2] \equiv \{q \mid L-a_1 \leqslant q \leqslant L+a_2\}$$

而对应的开区间以圆括号表示：

$$(L-a_1, L+a_2) \equiv \{q \mid L-a_1 < q < L+a_2\} \quad (6.4)$$

138 因此，[ ]与弱不等号 $\leqslant$ 相联系，而( )则与严格不等号 $<$ 相联系。但是在两种类型的区间中，较小的数 $(L-a_1)$ 总是列在前面。以后，我们还会遇到如(3,5]和[6,∞)这样的半开区间和半闭区间，其含义如下：

$$(3,5] \equiv \{x \mid 3 < x \leqslant 5\} \qquad [6,\infty) \equiv \{x \mid 6 \leqslant x < \infty\}$$

现在我们可以将 $L$ 的邻域定义为像(6.4)所定义的开区间，且此区间“包含”数 $L$。[①] 根据任意数 $a_1$ 和 $a_2$ 的大小，可为已知数 $L$ 构建各种不同的邻域。运用邻域的概念，函数的极限可定义如下：

> 当 $v$ 趋近于一个数 $N$，$q=g(v)$ 的极限是数 $L$，如果对 $L$ 的每个邻域，不管它多么小，总可以在函数的定义域中相应的 $N$ 的邻域(点 $v=N$ 除外)，使得 $N$ 邻域中每个 $N$ 值，其像落在 $L$ 的邻域内。

---

① 仅当我们考察直线(一维空间)上的点时，才可以将开区间视为点的邻域。在平面(二维空间)上一个点的情况下，其邻域必须被看作一个平面，比如环绕一个点的圆。

这个定义可以图 6.3 来加以解释。此图类似于图 6.2$a$。根据图6.2$a$ 我们已经学到的知识,我们知道图6.3 的$\lim\limits_{v \to N} q = L$。现在我们来证明 $L$ 确实满足极限的新定义。首先,选择 $L$ 的任意小的邻域,比如$(L - a_1, L + a_2)$。(应将其选择得更小一些,但为便于图形说明,我们将其相对放大了。现在我们构建 $N$ 的一个邻域,比如$(N - b_1, N + b_2)$,使得两个邻域共同定义一个矩形(图中阴影部分),其两个角位于给定的曲线上。这样便可以证明,对于 $N$ 的邻域中的每个 $v$ 值(不包括 $v = N$),对应的函数 $q = g(v)$的值位于选定的 $L$ 的邻域内。事实上,无论我们将 $L$ 的邻域选择得多么小,总可以找到具有上述性质的(相应的也很小)的 $N$ 邻域。因此,正如前面所证明的那样,$L$ 满足极限的定义。

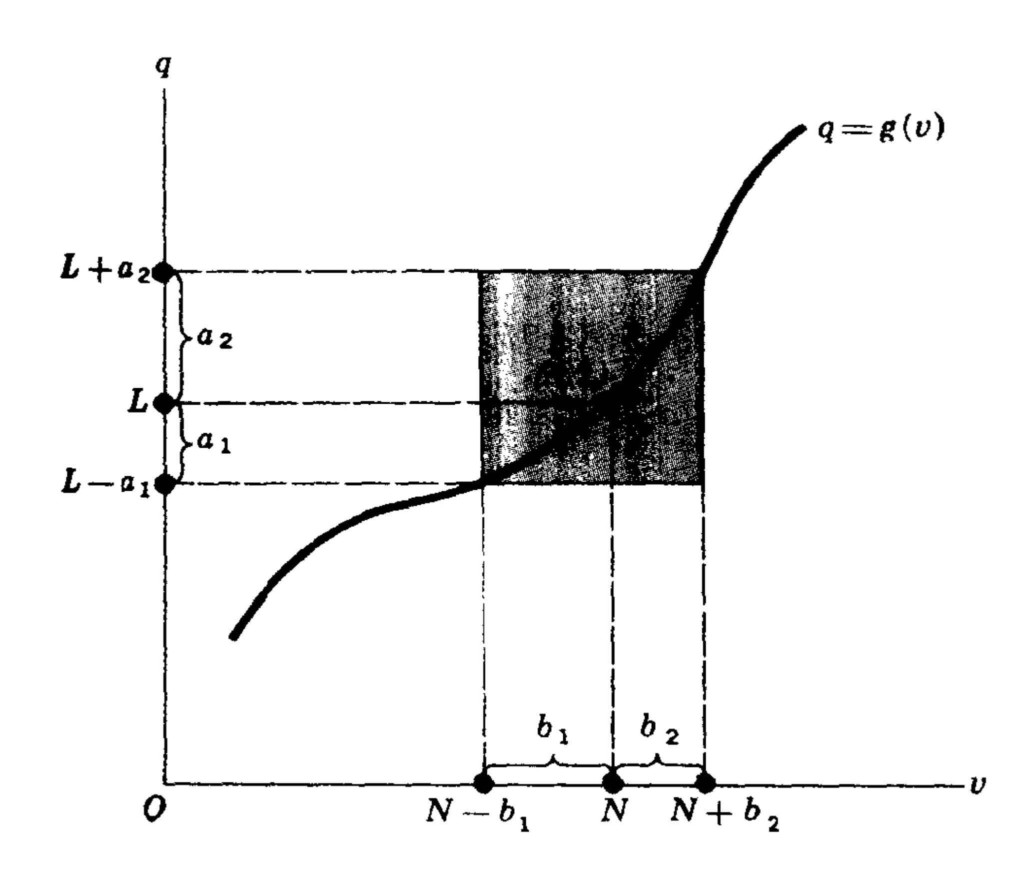



图 6.3

我们还可以将上述定义应用于图 6.2c 的阶段函数以证明无论$L_1$还是$L_2$，均不是极限$\lim\limits_{v\to N} q$。若我们选择$L_1$的一个极小的邻域，比如它仅在$L_1$两侧各有一根头发那么宽，那么，无论我们选择$N$的什么邻域，与两个邻域相联系的矩形不可能包容阶段函数中较低的部分。结果，对于$v>N$的任意值，相应的$q$值（位于阶段函数较低部分）将不会在$L_1$的邻域内，因此，$L_1$未能通过极限定义存在的检验。基于类似的理由，$L_2$也不是$\lim\limits_{v\to N} q$的候选值。事实上，在此情况下当$v\to N$时$q$不存在极限。

139 定义是否满足还可以通过代数方法而不是图形方法来检验。例如，再一次考察函数

$$q=\frac{1-v^2}{1-v}=1+v \qquad (v\neq 1) \tag{6.5}$$

在例 2 中已经知道$\lim\limits_{v\to 1} q=2$，因而这里有$N=1$和$L=2$。为验证$L=2$实际上是$q$的极限，我们必须证明，对于每个选定的$L$的邻域$(2-a_1,2+a_2)$，存在一个$N$的邻域$(1-b_1,\ 1+b_2)$，使得$v$无论何时在$N$的邻域内，$q$必然在所选定的$L$邻域内。实际上这意味着，对于给定的$a_1$和$a_2$值，无论它们多么小，总可以找到两个数$b_1$、$b_2$，使得不等式

$$1-b_1<v<1+b_2 \qquad (v\neq 1) \tag{6.6}$$

满足时，另一个不等式

$$2-a_1<q<2+a_2 \tag{6.7}$$

必然也成立。为找到这样一对数$b_1$和$b_2$，我们首先利用(6.5)式，将(6.7)式改写成

$$2-a_1<1+v<2+a_2 \tag{6.7'}$$

进而,可以将其变换成不等式

$$1 - a_1 < v < 1 + a_2 \tag{6.7''}$$

将(6.7)的变式(6.7″)与(6.6)式的比较表明,若我们选择两个数 $b_1$ 和 $b_2$,且使 $b_1 = a_1$, $b_2 = a_2$,两个不等式(6.6)和(6.7)总将同时成立。因此,$N$ 的邻域$(1 - b_1, 1 + b_2)$,如极限定义所需要的那样,在 $L = 2$ 时的确可以找到,这证明了 $L = 2$ 是 $q$ 的极限。 140

现在我们以相反的方式利用极限的定义,来证明另一个数,比如 3,不是函数(6.5)的极限$\lim\limits_{v \to 1} q$。假设 3 是此函数的极限,则对于 3 的每一个选定的邻域$(3 - a_1, 3 + a_2)$,存在一个 1 的邻域$(1 - b_1, 1 + b_2)$,使得当 $v$ 在 1 的邻域中时,$q$ 必定在 3 的邻域中。即无论何时不等式

$$1 - b_1 < v < 1 + b_2$$

满足,下面的另一不等式

$$3 - a_1 < 1 + v < 3 + a_2$$

或 $$2 - a_1 < v < 2 + a_2$$

必然也满足。要达到这一结果的唯一方式是选择 $b_1 = a_1 - 1$ 和 $b_2 = a_2 + 1$。这意味着 1 的邻域将是开区间$(2 - a_1, 2 + a_2)$。但按极限的定义,$a_1$ 和 $a_2$ 可取任意小的值,比如 $a_1 = a_2 = 0.1$。在此情况下,开区间为(1.9, 2.1),它完全位于横轴上的点 $v = 1$ 的右侧。因此,3 不满足极限的定义。还可以用类似的步骤证明除了 2 以外的任意数,均与本例中极限的定义相矛盾。

一般而言,若当 $v \to N$ 时,某一数满足 $q$ 的极限定义,则没有以外的数能满足定义。若极限存在,它将是唯一的。

## 练习 6.4

---

1 已知函数 $q=(v^2+v-56)/(v-7)$, $(v\neq7)$,求当 $v\to7$ 时 $q$ 的左极限和右极限。我们可以根据这些答案断定当 $v\to7$ 时,$q$ 有极限吗?

2 已知 $q=[(v+2)^3-8]/v$, $(v\neq0)$,求:

(a) $\lim_{v\to0} q$ (b) $\lim_{v\to2} q$ (c) $\lim_{v\to a} q$

3 已知 $q=5-1/v$, $(v\neq0)$, 求:

(a) $\lim_{v\to+\infty} q$ (b) $\lim_{v\to-\infty} q$

4 运用图 6.3 证明:我们不能将 $(L+a_2)$ 看作当 $v$ 趋近于 $N$ 时 $q$ 的极限。

---

141

## 6.5 关于不等式和绝对值的题外讨论

前面,我们已多次遇到不等号。在上一节最后的讨论中,我们还应用了不等式的数学运算。例如,在将(6.7′)变换成(6.7″)时,我们在不等式的每一边都同时减去 1。什么法则适于不等式(相对于方程而言)的运算呢?

### 不等式的运算法则

首先,我们描述不等式的一个重要性质:不等式是**可传递的**。即若 $a>b$, $b>c$, 则 $a>c$。因为等式(方程)也是可传递的,传递性既适用于“弱”不等式(⩾或⩽),也适用于“严格”不等式(>或<)。因此,我们有

$$a>b, b>c \Rightarrow a>c$$

$$a \geqslant b, b \geqslant c \Rightarrow a \geqslant c$$

这一性质使我们可以写出连不等式，如 $3<a<b<8$ 或 $7 \leqslant x \leqslant 24$ 等。(在书写连不等式时，不等式符号排成相同方向，通常将最小的数排在左边)。

不等式最重要的运算法则是不等式与数的加法、减法，以及不等式与数的乘法、除法，以及不等式的平方等。具体地说，这些法则如下：

**法则 I (加法与减法)**

$$a > b \Rightarrow a \pm k > b \pm k$$

若不等式两边加上或减去同样的数，不等式仍然成立。

**法则 II (乘法和除法)**

$$a > b \Rightarrow \begin{cases} ka > kb \ (k > 0) \\ ka < kb \ (k < 0) \end{cases}$$

不等式两边被一个正数乘，不等式方向不变；但被负的乘子乘则使不等式符号方向改变。

**例 1** 因 $6>5$，所以它被 3 乘得到 $3(6)>3(5)$ 或 $18>15$；但若被 $-3$ 乘则产生 $(-3)6<(-3)5$，或 $-18<-15$。

不等式被一个数 $n$ 除等价于它被一个数 $1/n$ 相乘，因而除法法则包含于乘法法则之中。

**法则 III (平方)** 142

$$a > b, (b \geqslant 0) \Rightarrow a^2 > b^2$$

若不等式两边均为非负，则当其两边取平方时不等式仍然成立。

**例 2** 因 $4>3$，且两边均为正，我们有 $4^2>3^2$，或 $16>9$。类似地，因 $2>0$，由此得到 $2^2>0^2$，或 $4>0$。

上述三个法则均按严格不等式形式表述，但若不等号“＞”被

弱不等号“$\geqslant$”替代，上述法则依然成立。

## 绝对值与不等式

当变量 $x$ 的定义域为开区间$(a, b)$时，此定义域可以集合$\{x \mid a<x<b\}$表示，或简单地以不等式 $a<x<b$ 表示。类似地，若定义域为闭区间$[a, b]$，则可以弱不等式 $a \leqslant x \leqslant b$ 表示。在区间的形式为$(-a, a)$这种特殊情况下，比如$(-10, 10)$，则此定义域既可以不等式 $-10<x<10$ 表示，或以不等式

$$|x| < 10$$

来表示。其中符号$|x|$表示 $x$ 的**绝对值**(或数值)。

对任意实数 $n$，$n$ 的绝对值定义如下：[①]

$$|n| \equiv \begin{cases} n \ (\text{若 } n > 0) \\ -n \ (\text{若 } n < 0) \\ 0 \ (\text{若 } n = 0) \end{cases} \tag{6.8}$$

注意，若 $n=15$，则$|15|=15$，但若 $n=-15$，则我们有

$$|-15| = -(-15) = 15$$

因而实际上任意实数的绝对值仅是去掉符号的数值。所以，我们总有$|n|=|-n|$。$n$ 的绝对值也称作 $n$ 的**模**。

给定表达式$|x|=10$，由(6.8)可知，$x$ 或者等于 10，或者等于$-10$。同理，表达式$|x|<10$ 意味着(1)若 $x>0$，则 $x \equiv |x| <10$，从而 $x$ 必小于 10；但(2)若 $x<0$，则依(6.8)，我们有 $-x \equiv |x| <10$，或 $x>-10$，从而 $x$ 必大于$-10$。因此，把这个结

① 绝对值的符号类似于一阶行列式的符号，但它们是完全不同的两个概念。一阶行列式的定义为$|a_{ij}| \equiv a_{ij}$，与 $a_{ij}$的正负号无关。按照绝对值$|n|$的定义，$n$ 的符号则会因之不同。在讨论的有关内容中，要弄清楚所考虑的是绝对值还是一阶行列式。

果的两部分结合起来，我们知道 $x$ 必定位于开区间 $(-10, 10)$ 内。一般而言，我们可以写成

$$|x| < n \Leftrightarrow -n < x < n \quad (n > 0) \tag{6.9}$$

143

它也可以拓展到如下弱不等式

$$|x| \leqslant n \Leftrightarrow -n \leqslant x \leqslant n \quad (n \geqslant 0) \tag{6.10}$$

因为绝对值本身是数，两个数 $m$、$n$ 的绝对值可以相加、减、乘、除。下述性质刻画了绝对值的特征：

$$|m| + |n| \geqslant |m + n|$$

$$|m| \cdot |n| = |m \cdot n|$$

$$\frac{|m|}{|n|} = \left|\frac{m}{n}\right|$$

令人感兴趣的是第一个性质，它涉及的是不等式而非方程。原因很简单：左边的表达式 $|m| + |n|$ 定义为两个数值（均视为正数）之和，表达式 $|m+n|$ 或者是两个数之和（若 $m$ 和 $n$ 均为正）或二者之差（若 $m$ 和 $n$ 的符号相反）。因而左侧的表达式可能大于右侧。

**例 3** 若 $m=5$, $n=3$, 则 $|m| + |n| = |m+n| = 8$。但若 $m=5$, $n=-3$, 则 $|m| + |n| = 5+3=8$，而

$$|m + n| = |5 - 3| = 2$$

则是较小的数。

另一方面，在其它两个性质中，$m$ 和 $n$ 的符号是相同还是相反是没有差别的。因为不等式右侧的积或商取绝对值，其符号无论如何，总会被去掉。

**例 4** 若 $m=7, n=8$，则 $|m| \cdot |n| = |m \cdot n| = 7(8) = 56$。但即使 $m=-7$ 和 $n=8$（符号相反），由下式仍可得到同样

的结果：

$$|m|\cdot|n| = |-7|\cdot|8| = 7(8) = 56$$

和

$$|m\cdot n| = |-7(8)| = 7(8) = 56$$

### 不等式的解

像方程一样，包含变量(比如 $x$)的不等式可能有解。若解存在，它将是一个使不等式成立的 $x$ 值的集合。这样的解本身一般是以不等式的形式表示。

**例 5**　求不等式的解

$$3x-3>x+1$$

像解方程一样，首先将含变量的项移至不等式一边。在不等式两边同时加 $(3-x)$，我们得到

144

$$3x-3+3-x>x+1+3-x$$

或　　$$2x>4$$

两边同乘 $\frac{1}{2}$(因 1/2 大于 0，所以不改变不等号的方向)则可得到解

$$x>2$$

它本身是一个不等式。这个解不是一个数，而是一个数集。因而我们也可以将其表示成集合 $\{x\mid x>2\}$ 或表示成开集 $(2,\infty)$。

**例 6**　解不等式 $|1-x|\leqslant 3$。首先，我们运用(6.10)式而不用绝对值的符号。这个已知不等式等价于

$$-3\leqslant 1-x\leqslant 3$$

或，在每一边减去 1 后，有

$$-4\leqslant -x\leqslant 2$$

将每一边乘 $-1$，我们则有

$$4 \geqslant x \geqslant -2$$

这里不等式的方向已颠倒了。先写出较小的一个数，我们可以将解表示成不等式的形式

$$-2 \leqslant x \leqslant 4$$

或表示成集合的形式 $\{x \mid -2 \leqslant x \leqslant 4\}$，或闭区间 $[-2,4]$。

有时，一个问题可能要求几个变量同时满足几个不等式，则我们必须解一个联合不等式组。例如本书的最后一篇，数学规划中就提出了这样的问题。

## 练习 6.5

---

1 解下列不等式：

(a) $3x-1<7x+2$　　(c) $5x+1<x+3$

(b) $2x+5<x-4$　　(d) $2x-1<6x+5$

2 若 $7x-3<0$ 和 $7x>0$，以连不等式对其进行表达，并求出其解。

3 解如下不等式：

(a) $|x+1|<6$　　(b) $|4-3x|<2$　　(c) $|2x+3|\leqslant 5$

---

# 6.6 极限定理

145

对变化率的兴趣使我们考察了导数的概念，而导数的差商极限的性质又激发我们研究极限的存在和计算问题。极限的基本计算步骤，如 6.4 节所述，包括令变量 $v$ 趋近于特定的数（比如 $N$），

以及观测 $q$ 所趋近的值等。但当实际计算一个函数的极限时，我们可以引用几个现成的极限定理，它可以极大地简化计算过程，对复杂的函数，尤其如此。

## 单一函数的极限定理

当仅涉及单个函数 $q=g(v)$时，可应用下述定理。

**定理 I** 若 $q=av+b$，则$\lim\limits_{v\to N}q=aN+b$($a$、$b$ 为常数)。

**例 1** 已知 $q=5v+7$，则我们有$\lim\limits_{v\to 2}q=5(2)+7=17$。类似地，$\lim\limits_{v\to 0}q=5(0)+7=7$。

**定理 II** 若 $q=g(v)=b$，则$\lim\limits_{v\to N}q=b$。

此定理表明，常函数的极限是该函数中的常数。它仅是定理 I 当 $a=0$ 时的一个特例。(读者早已在练习 6.2－3 中遇到了这一例子。)

**定理 III** 若 $q=v$，则$\lim\limits_{v\to N}q=N$。

若 $q=v^k$，则$\lim\limits_{v\to N}q=N^k$。

**例 2** 给定 $q=v^3$，我们有$\lim\limits_{v\to 2}q=(2)^3=8$。

读者可能已经注意到，在上述三个定理中，求当 $v\to N$ 时 $q$ 的极限，实际上是令 $v=N$。但这是一些特例，且它们并不损害"$v\to N$"并不意味着"$v=N$"的一般法则。

## 涉及两个函数的极限定理

若我们有两个具有同样自变量 $v$ 的函数 $q_1=g(v)$和 $q_2=h(v)$，且两个函数都有如下极限

$$\lim_{v \to N} q_1 = L_1 \qquad \lim_{v \to N} q_2 = L_2$$

其中 $L_1$ 和 $L_2$ 为两个有限的数,可应用下述定理:

**定理 IV (和—差极限定理)**

146

$$\lim_{v \to N}(q_1 \pm q_2) = L_1 \pm L_2$$

两个函数的和(差)的极限分别是其极限的和(差)。

特别地,我们要注意

$$\lim_{v \to N} 2q_1 = \lim_{v \to N}(q_1 + q_1) = L_1 + L_1 = 2L_1$$

它与定理 I 是一致的。

**定理 V (积的极限定理)**

$$\lim_{v \to N}(q_1 q_2) = L_1 L_2$$

两函数积的极限为其极限之积。

将其应用于函数的平方,给出

$$\lim_{v \to N}(q_1 q_1) = L_1 L_1 = L_1^2$$

这与定理 III 是一致的。

**定理 VI (商的极限定理)**

$$\lim_{v \to N} \frac{q_1}{q_2} = \frac{L_1}{L_2} \qquad (L_2) \neq 0$$

两个函数商的极限是其极限的商。自然,极限 $L_2$ 为非零,否则商无意义。

**例 3** 求 $\lim_{v \to 0}(1 + v)/(2 + v)$。因为我们这里有 $\lim_{v \to 0}(1 + v) = 1$ 和 $\lim_{v \to 0}(2 + v) = 2$,所求的极限为 1/2。

注意,$L_1$ 和 $L_2$ 代表有限的数,否则不能应用这些定理。在定理 VI 情况下,$L_2$ 也必须是非零的。若这些条件不满足,我们必须求助于 6.4 节中例 2 和例 3 所述的计算极限的方法,它们分

别与 $L_2$ 等于零，以及 $L_2$ 为无穷大的情况有关。

## 多项式函数的极限

掌握上述极限定理后，我们便可以轻松地计算任何多项式函数当 $v$ 趋于数 $N$ 时的极限了：

$$q = g(v) = a_0 + a_1 v + a_2 v^2 + \cdots + a_n v^n \tag{6.11}$$

因为其中的每一项的极限分别为

147 $\lim_{v \to N} a_0 = a_0 \qquad \lim_{v \to N} a_1 v = a_1 N \qquad \lim_{v \to N} a_2 v^2 = a_2 N^2$（等）

所以多项式函数的极限为（根据和极限定理）：

$$\lim_{v \to N} q = a_0 + a_1 N + a_2 N^2 + \cdots + a_n N^n \tag{6.12}$$

我们注意到，此极限实际上也等于 $g(N)$，即等于当 $v = N$ 时(6.11)的函数值。这个特别的结果将证明讨论多项式函数的连续性的重要性。

## 练习 6.6

1 求函数 $q = 8 - 9v + v^2$ 的极限：

(a) 当 $v \to 0$ (b) 当 $v \to 3$ (c) 当 $v \to -1$

2 求 $q = (v+2)(v-3)$ 的极限：

(a) 当 $v \to -1$ (b) 当 $v \to 0$ (c) 当 $v \to 4$

3 求 $q = (3v+5)/(v+2)$ 的极限：

(a) 当 $v \to 0$ (b) 当 $v \to 5$ (c) 当 $v \to -1$

## 6.7 函数的连续性与可微性

现在我们运用前面对极限概念和极限计算的讨论来定义函数的连续性和可微性。这些概念与我们感兴趣的函数的导数直接相关。

### 函数的连续性

当 $v$ 趋近于定义域中的点 $A$ 时，函数 $q=g(v)$ 的极限存在，且等于 $g(N)$，即等于函数在 $v=N$ 时的值，则我们说函数在点 $N$ 是**连续的**。如上所述，**连续性**这个概念至少有以下三个要求：(1) 点 $N$ 必须在函数的定义域内，即 $g(N)$ 有定义；(2) 当 $v\to N$ 时，函数的极限存在，即 $\lim\limits_{v\to N} g(v)$ 存在；(3) 极值必等于 $g(N)$ 的值，即 $\lim\limits_{v\to N} g(v)=g(N)$。

注意到这一点是重要的：尽管在图 6.3 讨论曲线的极限时，点 $(N,L)$ 未被考虑，但在这里我们必须将其考虑在内。特别是，正如上述第三条要求所具体说明的那样，在将函数视为于点 $N$ 连续以前，点 $(N,L)$ 必定在函数的图形上。

我们来检验图 6.2 所示的函数是否是连续的。在图形 $A$ 中，在点 $N$，三条要求均被满足：点 $N$ 在定义域内；当 $v\to N$ 时，$q$ 具 148 有极值 $L$；极值 $L$ 恰好等于函数在点 $N$ 的值。因此，该曲线所表示的函数在 $N$ 处是连续的。同样，图 6.2$b$ 所描述的函数也是连续的，因为 $L$ 是当 $v$ 趋近于定义域内 $N$ 点时函数的极值，且 $L$ 也是在点 $N$ 时的函数值。后一图例足以证明，函数在点 $N$ 的连续

性并一定要求函数的图形在 $v=N$ 处是必须“平滑的”。因为在图 6.2$b$ 中点 $(N,L)$ 实际上是一个“尖”折点，但函数在此 $v$ 值处仍是连续的。

当函数 $q=g(v)$ 在区间 $(a,b)$ 中所有的 $v$ 值处均连续时，则此函数在这一区间内连续。若函数在定义域内的子集 $S$(其中子集 $S$ 可能是几个不相交的区间的并集)中的所有点连续，则我们称函数在 $S$ 中连续。最后，若函数在其定义域中的所有点连续，则称函数在其定义域中连续，但甚至在后一种情况下，对于某些不在定义域中的点，比如 $v=5$，函数的图形仍然会表现出不连续性(缺口)。

再回到图 6.2，我们看到图($c$)中函数在点 $N$ 处是不连续的，因为极限在该点不存在。与连续性的第二条要求相违背。但是在区间 $(0,N)$ 和 $[N,\infty)$，函数确实满足连续性的三条要求。图 $d$ 在 $v=N$ 处明显也是不连续的。但这一次，不连续性是因为点 $N$ 不在定义域内，违反连续性的第一条要求。

由图 6.2 可以看出，连续性与角点是相容的，如图 $b$ 所示；但像图 $c$ 和 $d$ 中的缺口是不允许的。情况的确如此。因此，粗略说来，如果在给定区间中作图时不用抬起笔，就可以说函数在该区间中是连续的，即使出现角点，也是允许的，但不允许有间断的情况出现。

## 多项式函数和有理函数

我们现在考察某些常见函数的连续性。对于任意多项式函数，如(6.11)中的 $q=g(v)$，由(6.12)我们已经知道 $\lim\limits_{v\to N} q$ 存在，且等于函数在 $N$ 点时的值。因为 $N$ 是函数定义域中的一个点

(任意点),我们可以得出这一结论:任何多项式函数在其定义域中是连续的。这是一条非常重要的信息,因为多项式函数是经常遇到的。

那么,有理函数又如何呢? 关于其连续性,存在一个很有价值的定理。该定理表明:在定义域内连续的任意有限个函数的和、差、积、商,在定义域内同样分别是连续的。因此,任意有理函数(两个多项式函数的商)在定义域内必然也是连续的。

**例1** 有理函数



$$q = g(v) = \frac{4v^2}{v^2 + 1}$$

对全部有限实数有定义。因此其定义域为$(-\infty, +\infty)$。对于定义域中任意数 $N$,由商极限定理,$q$ 的极限为

$$\lim_{v \to N} q = \frac{\lim\limits_{v \to N}(4v^2)}{\lim\limits_{v \to N}(v^2 + 1)} = \frac{4N^2}{N^2 + 1}$$

它等于 $g(N)$。因此在 $N$,连续性的三条要求均满足,所以函数在定义域中是连续的。

**例2** 有理函数

$$q = \frac{v^3 + v^2 - 4v - 4}{v^2 - 4}$$

在 $v=2$ 和 $v=-2$ 无意义。因 $v$ 的这两个值不在定义域中,函数在 $v=2$ 和 $v=-2$ 是不连续的,尽管当 $v \to 2$ 和 $v \to -2$ 时,$q$ 的极限是存在的。从图形上看,函数在 $v$ 的这两个值处会出现两个缺口。但对于在定义域的 $v$ 的其它值,函数是连续的。

## 函数的可微性

前面的讨论已为我们提供了一些基本的工具,用以判断当自

变量趋近于某个选定的值 $x_0$ 时，函数的极限是否存在。因此，我们可以尝试取当 $x$ 趋近于某个选定的值 $x_0$，函数 $y=f(x)$ 的极限。但我们还可以在不同水平上应用“极限”这一概念，并在当 $\Delta x \to 0$ 时，取函数的差商 $\Delta y/\Delta x$ 的极限。在这两种不同水平上取极限的结果与函数 $f$ 的两种不同却相联系的性质有关。

取函数 $y=f(x)$ 的极限，与上一小节的讨论一致，我们可以检验函数 $f$ 在 $x=x_0$ 处是否连续。连续性的条件为：(1) $x=x_0$ 必定在函数的定义域内；(2) 当 $x \to x_0$ 时，$y$ 的极限存在；(3) 所求的极限必等于 $f(x_0)$。当这些条件满足，我们可以写成

$$\lim_{x \to x_0} f(x) = f(x_0) \qquad [\text{连续性条件}]$$

另一方面，当我们对差商 $\Delta y/\Delta x$ 应用“极限”的概念时，我们便在研究函数 $f$ 在 $x=x_0$ 是否可微，即导数 $dy/dx$ 在 $x=x_0$ 是否存在，或 $f'(x_0)$ 是否存在的问题了。这里使用“可微的”这一术语是因为求得导数 $dy/dx$ 的过程被看作微分（也称求导）的过程。150 当且仅当 $\Delta x \to 0$，$\Delta y/\Delta x$ 在 $x=x_0$ 的极限存在时，以符号表示函数 $f$ 的可微性：

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \tag{6.14}$$

$$\equiv \lim_{\Delta x \to 0} \frac{f(x_0+\Delta x) - f(x_0)}{\Delta x} \qquad [\text{可微性条件}]$$

连续性和可微性这两个性质，是紧密相关的：函数的连续性是可微性的必要条件（但我们后面将看到，它不是充分条件）。这意味着，函数要在 $x=x_0$ 处可微，必须首先通过在 $x=x_0$ 点的连续性检验。为证明这一点，我们需要证明，已知函数 $y=f(x)$，由它在 $x=x_0$ 处可微，便可以推定它在 $x=x_0$ 处连续，即满足条件

(6.13)便可满足条件(6.14)。但在证明之前，我们需简化一下符号：(1)以符号 $N$ 代替 $x_0$；(2)以符号 $x$ 代替 $(x_0+\Delta x)$——这种替代是合理的。因为 $x$ 变化后的值可以是任何数(取决于变化的大小)，因此可以是一个以 $x$ 表示的变量。两种符号的等价性在图 6.4 中已经表示清楚了，其中原符号与新符号一起标出。注意，由于符号的变化，$\Delta x$ 现在变成 $(x-N)$，从而"$\Delta x\to 0$"变成"$x\to N$"。它类似于与函数 $q=g(v)$ 相联系的"$v\to N$"。相应的，151

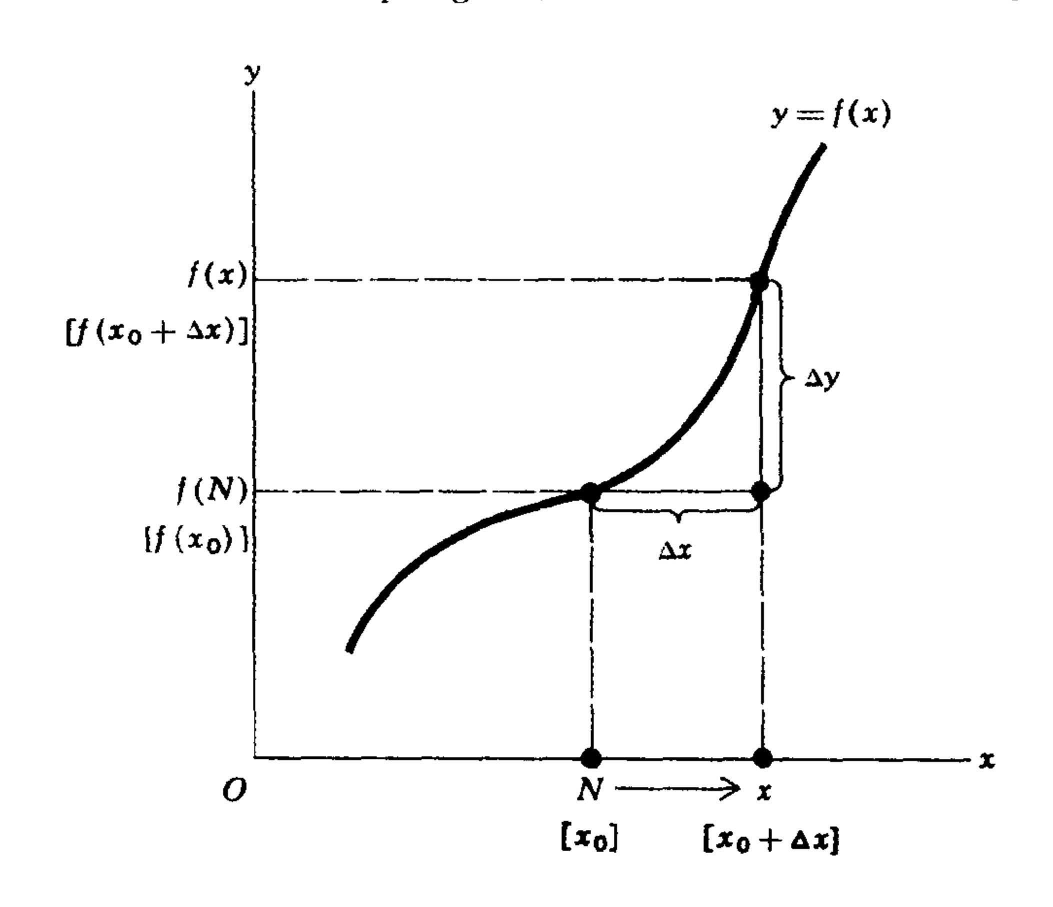



图 6.4

(6.13)和(6.14) 现在可分别改写为

$$\lim_{x\to N} f(x) = f(N) \tag{6.13'}$$

$$f'(N) = \lim_{x\to N}\frac{f(x)-f(N)}{x-N} \tag{6.14'}$$

因此，我们想证明的是，由可微性条件(6.14′)可推出连续性

条件(6.13′)。首先,因为记号 $x \to N$ 意味着 $x \neq N$,所以 $x - N$ 不等于零,可以写出下列恒等式

$$f(x) - f(N) \equiv \frac{f(x) - f(N)}{x - N}(x - N) \qquad (6.15)$$

当 $x \to N$ 时对(6.15)每一边取极限,得到下面的结果:

$$\begin{aligned}
\text{左边} &= \lim_{x \to N} f(x) - \lim_{x \to N} f(N) && [\text{差的极限定理}] \\
&= \lim_{x \to N} f(x) - f(N) && [f(N)\ \text{是常数}] \\
\text{右边} &= \lim_{x \to N} \frac{f(x) - f(N)}{x - N} \lim_{x \to N}(x - N) && [\text{积的极限定理}] \\
&= f'(N)(\lim_{x \to N} x - \lim_{x \to N} N) && [\text{由}(6.14')\ \text{和差的极限定理}] \\
&= f'(N)(N - N) = 0
\end{aligned}$$

注意,如果条件(6.14′)不成立,我们就不能写出上述结果。因为(6.15)右边的表达式(以及左边的表达式)没有极限。但若 $f'(N)$ 确实存在,则如上所述,上式左右两边的极限存在。而且,当左、右两边的极限相等时,我们得到 $\lim\limits_{x \to N} f(x) - f(N) = 0$,这与(6.13′)是一致的。因而我们便由(6.13)所表示的可微性,推导出了(6.14)所表示的连续性。一般而言,若函数在其定义域中每点均可微,则我们便可断定函数在定义域中连续。

尽管可微性意味着连续性,但反之未必成立。即连续性是可微性的**必要**条件,但不是**充分**条件,为证明这一点,我们仅需举一反例。我们考察函数:

$$y = f(x) = |x - 2| + 1 \qquad (6.16)$$

它的图形如图 6.5 所示。容易证明,此函数当 $x = 2$ 时尽管连续,但不可微。函数在 $x = 2$ 处连续是很易证明的。首先,$x = 2$ 在函

数定义域内;其次,当 $x \to 2$ 时 $y$ 的极限存在;具体而言,$\lim_{x\to 2^+} = \lim_{x\to 2^-} y = 1$。第三,也可求得 $f(2)=1$。因此连续性的三个条件均满足。为证明函数 $f$ 在 $x=2$ 不可微,我们必须证明差商的极限

$$\lim_{x\to 2}\frac{f(x)-f(2)}{x-2}=\lim_{x\to 2}\frac{|x-2|+1-1}{x-2}=\lim_{x\to 2}\frac{|x-2|}{x-2}$$

不存在。这就要证明左极限和右极限不相等。因为考察右极限可知,$x$ 必大于 2,按照绝对值定义(6.8),我们有 $|x-2|=x-2$。因此,右极限为

$$\lim_{x\to 2^+}\frac{|x-2|}{x-2}=\lim_{x\to 2^+}\frac{x-2}{x-2}=\lim_{x\to 2^+}1=1$$

另一方面,在考察左极限时,$x$ 必定小于 2;因此,按照(6.8),$|x-2|=-(x-2)$。结果,左极限为

$$\lim_{x\to 2^-}\frac{|x-2|}{x-2}=\lim_{x\to 2^-}\frac{-(x-2)}{x-2}=\lim_{x\to 2^-}(-1)=-1$$

它不同于右极限。这就证明了连续性不能保证可微性。简言之,所有可微的函数必定是连续的,但不是所有连续的函数都是可微的。

在图 6.5 中,函数在 $x=2$ 处显然不可微,因为在点(2,1)没有有定义的切线,因而在此点也就没有明确的斜率。具体而言,在该点左侧,曲线的斜率为 −1,而右侧的斜率为 +1,在 $x=2$ 处,左右两边的斜率并没有趋向相同数值的倾向。当然,点(2,1)是一个特殊的点;它是曲线上的唯一角点。在曲线上其它的点,导数有定义,函数可微。更具体地说,(6.16)中的函数可以分成如下两个线性函数: 153

左侧: $y=-(x-2)+1=3-x \quad (x\leqslant 2)$

右侧: $y=(x-2)+1=x-1 \quad (x>2)$

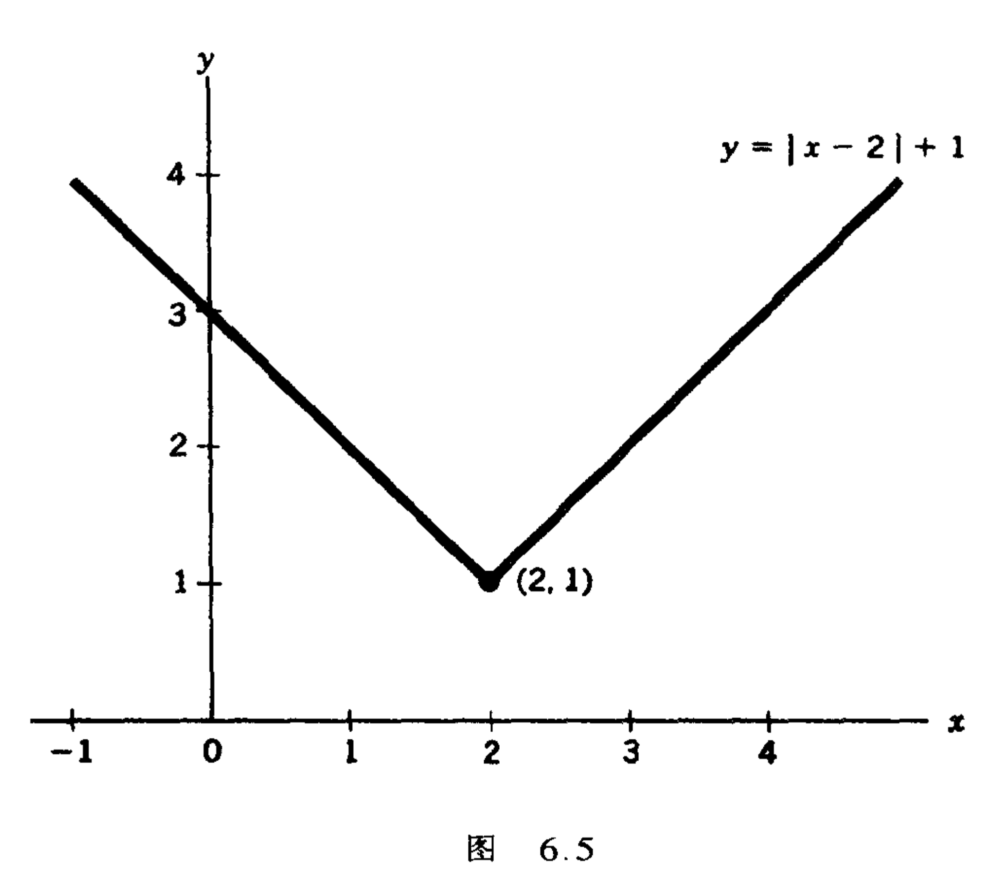



图 6.5

左侧在定义域区间$(-\infty,2)$是可微的，右侧在定义域区间$(2,\infty)$是可微的。

一般而言，可微性的条件比连续性更严格，因为除了连续性的条件之外，可微性还要求额外的条件。连续性仅排除了缺口的存在，而可微性还要求不存在角点。因此，可微性不仅要求函数具有连续性，而且还要求函数(曲线)具有“平滑性”。经济学中所遇到的大多数具体的函数均具有处处可微的性质。所以，我们一般假定所运用的一般函数是处处可微的。我们在以后的讨论中也将如此。

## 练习 6.7

1 函数 $y=f(x)$在 $x=x_0$ 处连续的三个要求中，只要有一条违背了，

函数 $y=f(x)$在 $x=x_0$ 处便不连续。请构造三种图形分别说明违背上述每一条要求的情形。

2　将所有有限实数集作为函数 $q=g(v)=v^2-7v-3$:

($a$) 求$\lim\limits_{v \to N} q$($N$ 为有限实数)。

($b$) 检验极限是否等于 $g(N)$。

($c$) 检验函数 $g(v)$在 $N$ 点及其定义域$(-\infty,+\infty)$的连续性。

3　给定函数 $q=g(v)=\dfrac{v+2}{v^2+2}$:

($a$) 运用极限定理求$\lim\limits_{v \to N} q$,$N$ 为有限实数。

($b$) 检验极限是否等于 $g(N)$。

($c$) 检验函数 $g(v)$在点 $N$ 以及它的定义域$(-\infty+\infty)$内是否连续。

4　已知 $y=f(x)=\dfrac{x^2-x-12}{x-2}$:

($a$) 能否应用商极限定理求 $x \to 4$ 时此函数的极限?

($b$) 函数在 $x=4$ 处连续吗? 为什么?

($c$) 求一个与上述函数等价的函数(对于 $x \neq 4$),并由此等价函数求原函数 $y$ 当$x \to 4$ 时的极限。

5　在例 2 的有理函数中,分子可被分母整除,其商为$(v+1)$。我们能 154 因此而以 $q=v+1$ 代替该函数吗? 为什么能? 为什么不能?

6　根据图 2.8 中 6 个函数的图形,你能推断这些函数在其定义域的每一点都可微吗? 请予以解释。

---

155

# 第7章 微分法则及其在比较静态学中的应用

比较静态分析的核心问题,即求变化率的问题,仅当所考察的 $x$ 变化很小时,才与求函数 $y=f(x)$的导数的问题相一致。尽管导数 $dy/dx$ 被定义为当 $v\to0$ 时,$q=g(v)$的差商的极限,但每次求导数时,并没有必要都进行取极限的过程,因为存在各种微分(求导)法则可使我们直接求得所需要的导数。因此,我们并不立即开始涉及比较静态模型,而是首先学习一些微分法则。

## 7.1 一元函数的微分法则

首先,我们讨论分别应用于以下三种类型的一元函数的微分法则:$y=k$(常值函数),$y=x^n$ 以及 $y=cx^n$(幂函数)。这些函数均有平滑、连续曲线,因而是处处可微的。

### 常值函数微分法则

常值函数 $y=f(x)=k$ 的导数等于零,即对所有的 $x$ 值,其导数均为零。可用符号表示如下:

$$\frac{dy}{dx}=0 \qquad 或 \qquad \frac{dk}{dx}=0 \qquad 或 \qquad f'(x)=0$$

156 事实上,我们还可以写成如下形式:

$$\frac{d}{dx}y = \frac{d}{dx}f(x) = \frac{d}{dx}k = 0$$

其中导数符号被分成两部分，$d/dx$ 为一部分，$y$(或 $f(x)$或 $k$)为另一部分。第一部分 $d/dx$ 可视为算子符，它向我们说明一种特定的数学运算。正如算子符 $\sqrt{\quad}$ 可表明取平方根一样，算子符 $d/dx$ 表示对变量 $x$ 取导数，或者对 $x$ 微分。被运算或被微分的函数在第二部分中标示出来，在这里为 $y = f(x) = k$。

这个微分法则的证明如下：给定 $f(x) = k$，对于任意 $N$ 值，我们有 $f(N) = k$。因此 $f'(N)$的值，即在 $x = N$ 处的导数值，按定义(6.13)应为

$$f'(N) = \lim_{x \to N} \frac{f(x) - f(N)}{x - N} = \lim_{x \to N} \frac{k - k}{x - N} = \lim_{x \to N} 0 = 0$$

进而，因 $N$ 表示 $x$ 的任意值，结果 $f'(N) = 0$ 可立即一般化为 $f'(x) = 0$，法则得证。

将 $f'(x) = 0$ 与看起来类似但实际不同的 $f'(x_0) = 0$ 作明确区分是极为重要的。由 $f'(x) = 0$，我们知道导函数 $f'$ 对所有 $x$ 值均为零。但是写成 $f'(x_0) = 0$，我们只能将导数为零与一个特定的 $x$ 值，即 $x = x_0$，联系起来。

如前所述，函数的导数的几何表示是曲线的斜率，常值函数，比如固定成本函数 $C_F = f(Q) = 1200$ 美元的图形，是一条斜率始终为零的水平线。相应地，对所有 $Q$ 值，导数必然始终为零：

$$\frac{d}{dQ}C_F = \frac{d}{dQ}1200 = 0 \qquad \text{或} \qquad f'(Q) = 0$$

## 幂函数微分法则

幂函数 $y = f(x) = x^n$ 的导数是 $nx^{n-1}$。用符号可表示成

$$\frac{d}{dx}x^n = nx^{n-1} \quad 或 \quad f'(x) = nx^{n-1} \tag{7.1}$$

**例 1** $y = x^3$ 的导数为 $\frac{dy}{dx} = \frac{d}{dx}x^3 = 3x^2$.

**例 2** $y = x^9$ 的导数为 $\frac{d}{dx}x^9 = 9x^8$.

这一法则对 $x$ 的任意实数次幂都是成立的,即指数可以是任意实数。但我们仅在 $n$ 为正整数的情况下证明这一法则。 157

在最简单的情况下,即 $n = 1$ 时,函数为 $f(x) = x$,按照幂函数微分法则,导数为

$$f'(x) = \frac{d}{dx}x = 1(x^0) = 1$$

这个结果由(6.14′)中 $f'(N)$的定义很容易得到证明。给定 $f(x) = x$,任意 $x$ 值,比如 $x = N$ 的导数值为

$$f'(N) = \lim_{x \to N}\frac{f(x) - f(N)}{x - N} = \lim_{x \to N}\frac{x - N}{x - N} = \lim_{x \to N}1 = 1$$

因 $N$ 表示 $x$ 的任意值,所以可以写成 $f'(x) = 1$。这证明了在 $n = 1$ 时,法则成立。将此结果用几何图形表示出来,我们看到函数 $y = f(x) = x$ 画成一条 45°的直线,其斜率始终为 1。

对于大于 1 的整数,$n = 2,3\cdots$的情况,我们首先注意下列等式

$$\frac{x^2 - N^2}{x - N} = x + N \qquad [右边两项]$$

$$\frac{x^3 - N^3}{x - N} = x^2 + Nx + N^2 \qquad [右边三项]$$

$$\vdots$$

$$\frac{x^n - N^n}{x - N} = x^{n-1} + Nx^{n-2} + N^2x^{n-3} + \cdots + N^{n-1} \tag{7.2}$$

[右边 $n$ 项]

根据(7.2),我们可以将幂函数 $f(x)=x^n$ 在 $x=N$ 的导数表示如下:

$$
\begin{aligned}
f'(N) &= \lim_{x \to N} \frac{f(x)-f(N)}{x-N} = \lim_{x \to N} \frac{x^n - N^n}{x-N} \\
&= \lim_{x \to N}(x^{n-1} + Nx^{n-2} + \cdots + N^{n-1}) \quad [\text{by}(7.2)] \\
&= \lim_{x \to N} x^{n-1} + \lim_{x \to N} Nx^{n-2} + \cdots + \lim_{x \to N} N^{n-1} \quad [\text{和的极限定理}] \\
&= N^{n-1} + N^{n-1} + \cdots + N^{n-1} \quad [n\text{ 项的总和}] \\
&= nN^{n-1} \qquad (7.3)
\end{aligned}
$$

同样,$N$ 为任意 $x$ 值,所以最后结果可推广为

$$f'(x) = nx^{n-1}$$

这证明了对任意正整数 $n$,法则成立。

正如前面所提到的那样,即使幂函数 $x^n$ 的指数 $n$ 不是正整数,仍可应用这一法则。下面的几个例子可以说明这种应用。

**例 3** 求 $y=x^0$ 的导数。应用(7.1),我们求得 158

$$\frac{d}{dx}x^0 = 0(x^{-1}) = 0$$

**例 4** 求 $y=1/x^3$ 的导数。此式含有幂数的倒数,但我们将函数重写成 $y=x^{-3}$,我们可再应用(7.1)得到导数:

$$\frac{d}{dx}x^{-3} = -3x^{-4} \quad \left[= \frac{-3}{x^4}\right]$$

**例 5** 求 $y=\sqrt{x}$ 的导数。此式包含平方根,但因$\sqrt{x}=x^{1/2}$,可求出导数如下:

$$\frac{d}{dx}x^{1/2} = \frac{1}{2}x^{-1/2} \qquad \left[= \frac{1}{2\sqrt{x}}\right]$$

导数本身是自变量 $x$ 的函数。如在例 1 中，导数为 $dy/dx = 3x^2$，或 $f'(x) = 3x^2$，因而不同的 $x$ 值，将产生不同的导数值，如

$$f'(1) = 3(1)^2 = 3 \qquad f'(2) = 3(2)^2 = 12$$

这些具体的导数值还可以表示成

$$\left.\frac{dy}{dx}\right|_{x=1} = 3 \qquad \left.\frac{dy}{dx}\right|_{x=2} = 12$$

显然记号 $f'(1)$ 和 $f'(2)$ 更受欢迎，因为它们更简洁。

认识到要求导数值 $f'(1)$，$f'(2)$ 等，必须首先对函数 $f(x)$ 微分，以得到导函数 $f'(x)$，然后再令 $f'(x)$ 中的 $x$ 为具体的值，是至关重要的。在微分前将具体值代入原函数 $f(x)$ 是不允许的。我们举例说明一下，若我们在微分前令例 1 中的 $x = 1$，函数将退化为 $y = x = 1$，一个常值函数，其导数为零，而非正确的答案 $f'(x) = 3x^2$。

## 幂函数微分法则的一般化

当一个常数乘子 $c$ 出现于幂函数中，从而有 $f(x) = cx^n$，其导数为

$$\frac{d}{dx}cx^n = cnx^{n-1} \qquad \text{或} \qquad f'(x) = cnx^{n-1}$$

这个结果表明，在求 $cx^n$ 的微分时，我们仅需保持常数乘子 $c$ 不变，按(7.1)对 $x^n$ 项求微分。

159

**例 6** 给定 $y = 2x$，我们有 $dy/dx = 2x^0 = 2$。

**例 7** 给定 $f(x) = 4x^3$，导数为 $f'(x) = 12x^2$。

**例 8** $f(x) = 3x^{-2}$ 的导数为 $f'(x) = -6x^{-3}$。

对于这个新法则的证明，我们考察下述事实：对任意 $x$ 值，比

如 $x=N$, $f(x)=cx^n$ 的导数值是

$$f'(N)=\lim_{x\to N}\frac{f(x)-f(N)}{x-N}=\lim_{x\to N}\frac{cx^n-cN^n}{x-N}=\lim_{x\to N}c\left(\frac{x^n-N^n}{x-N}\right)$$

$$=\lim_{x\to N}c\lim_{x\to N}\frac{x^n-N^n}{x-N} \qquad [\text{积的极限定理}]$$

$$=c\lim_{x\to N}\frac{x^n-N^n}{x-N} \qquad [\text{常数的极限}]$$

$$=cnN^{n-1} \qquad [\text{由}(7.3)]$$

由于 $N$ 为 $x$ 的任意值,最后的结果可立即一般化为 $f'(x)=cnx^{n-1}$,法则得证。

## 练习 7.1

---

1 求下述每一函数的导数:

$(a)y=x^{13}$ $(c)y=7x^6$ $(e)w=-4u^{1/2}$

$(b)y=63$ $(d)w=3u^{-1}$

2 求下述函数的导数:

$(a)\frac{d}{dx}(-x^{-4})$ $(c)\frac{d}{dw}9w^4$ $(e)\frac{d}{du}au^b$

$(b)\frac{d}{dx}7x^{1/3}$ $(d)\frac{d}{dx}cx^2$

3 求下述每一函数的 $f'(1)$、$f'(2)$的值:

$(a)\ y=f(x)=18x$ $(c)f(x)=-5x^{-2}$ $(e)f(w)=6w^{1/3}$

$(b)y=f(x)=cx^3$ $(d)f(x)=\frac{3}{4}x^{4/3}$

4 绘出使得导函数 $f'(x)=0$ 的原函数 $f(x)$的图形。然后再绘出以 $f'(x_0)=0$ 为特征的函数 $g(x)$的图形。

---

# 7.2 相同变量的两个或两个以上函数的微分法则

上节给出的三个法则均与单个给定函数 $f(x)$有关。现在假设我们有相同变量 $x$ 的两个可微函数，比如 $f(x),g(x)$，我们想160对由两个函数的和、差、积、商求微分。在此情况下，我们是否有合适的法则可应用？更具体地说，给定两个函数，比如 $f(x)=3x^2$ 和 $g(x)=9x^{12}$，我们如何求 $3x^2+9x^{12}$ 的导数呢？或者$(3x^2)(9x^{12})$的导数呢？

### 和、差的微分法则

两个函数和(差)的导数等于两个函数的导数的和(差)。

$$\frac{d}{dx}[f(x)\pm g(x)]=\frac{d}{dx}f(x)\pm\frac{d}{dx}g(x)=f'(x)\pm g'(x)$$

这个证明涉及到导数定义和各极限定理的应用。我们将省略对它的证明，而只验证其有效性并举例说明其应用。

**例1** 由函数 $y=14x^3$，我们得到导数 $dy/dx=42x^2$。但 $14x^3=5x^3+9x^3$，从而 $y$ 可以看作两个函数 $f(x)=5x^3$ 和 $g(x)=9x^3$ 的和。按照函数和的微分法则，则我们有

$$\frac{dy}{dx}=\frac{d}{dx}(5x^3+9x^3)=\frac{d}{dx}5x^3+\frac{d}{dx}9x^3$$
$$=15x^2+27x^2=42x^2$$

它与我们前面的结果是一致的。

这个按两函数表述的法则，可以很容易推广至更多的函数。因此，写出下式也是正确的

$$\frac{d}{dx}[f(x) \pm g(x) \pm h(x)] = f'(x) \pm g'(x) \pm h'(x)$$

**例 2** 引用例 1 的函数 $y = 14x^3$，可以将其写成 $y = 2x^3 + 13x^3 - x^3$。按照和差微分法则，这个式子的导数为

$$\frac{dy}{dx} = \frac{d}{dx}(2x^3 + 13x^3 - x^3) = 6x^2 + 39x^2 - 3x^2 = 42x^2$$

这再一次验证了上面的答案。

这一法则是极其重要的。有了它，就可以求出任意多项式的导数，因为多项式不过是幂函数的和。

**例 3** $$\frac{d}{dx}(ax^2 + bx + c) = 2ax + b$$

**例 4** $$\frac{d}{dx}(7x^4 + 2x^3 - 3x + 37) = 28x^3 + 6x^2 - 3 + 0$$ 161

$$= 28x^3 + 6x^2 - 3$$

注意，在最后两个例子中，常数 $c$ 和 37 对导数实际上不产生任何影响，因为常数项导数为零。与在微分时依然保留的乘子常数相反，加和常数在求导时去掉了。这一事实为下述众所周知的经济原理提供了数学解释：厂商的固定成本不影响边际成本。给定短期成本函数：

$$C = Q^3 - 4Q^2 + 10Q + 75$$

边际成本函数（产出变化无穷小）是差商 $\Delta C / \Delta Q$ 的极限，或成本函数的导数

$$\frac{dC}{dQ} = 3Q^2 - 8Q + 10$$

而固定成本以加和常数 75 表示。因为后者在求导过程中消失了，固定成本的大小显然不对边际成本产生影响。

一般而言，若原函数 $y = f(x)$ 表示总函数，则导函数 $dy/dx$

则是边际函数。当然，两个函数均可对变量 $x$ 作出图形。并且由于函数导数与其曲线斜率之间的对应性，对于每一个 $x$ 值，边际函数应表示出总函数在该 $x$ 值的斜率。在图 7.1$a$ 中，线性（常斜



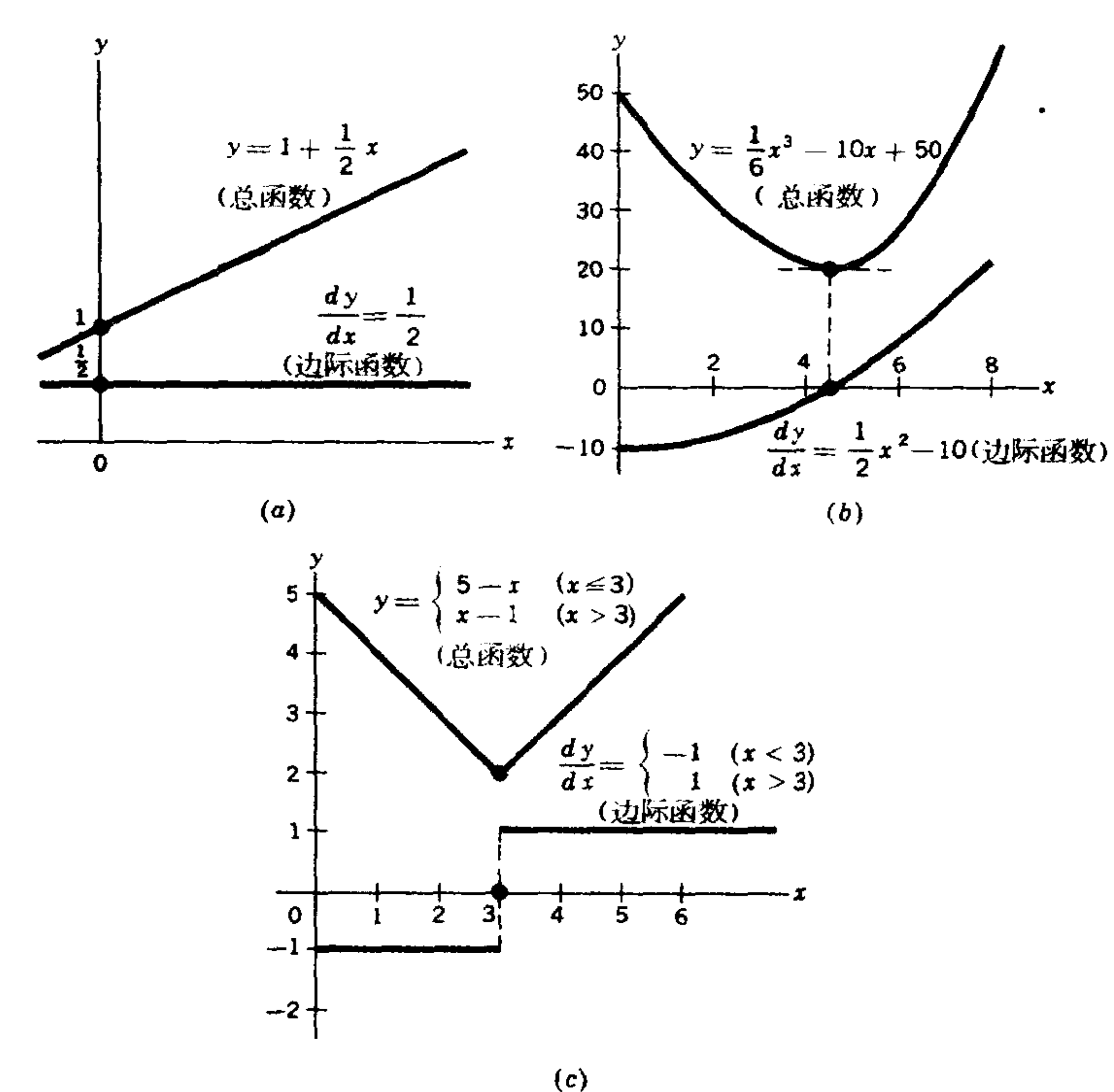



图 7.1

率）总函数看来具有常值边际函数。另一方面，图 7.1$b$ 中的非线性（斜率在变化）总成本导致弯曲的边际函数，当总函数斜率为负（正）时，曲线位于横轴的下（上）方。最后，读者可能由图 7.1$c$ 注意到，总函数的"不光滑性"会使边际函数或导函数曲线出现"缺口"（不连续）。它与图 7.1$b$ 处处平滑的总函数形成了鲜明的对

照，后者给出了连续的边际函数。因此，原函数的平滑性与导函数的连续性是联系在一起的。特别是，我们可以不说某个函数是处处平滑的（和可微的），而是称其具有连续导函数的特征，并称其为连续可微的函数。

## 积的微分法则

两个（可微分）函数的积的导数等于第一个函数乘以第二个函数的导数，加上第二个函数乘以第一个函数的导数：

$$\frac{d}{dx}[f(x)g(x)] = f(x)\frac{d}{dx}g(x) + g(x)\frac{d}{dx}f(x)$$

$$= f(x)g'(x) + g(x)f'(x) \qquad (7.4)$$

**例 5** 求函数 $y=(2x+3)(3x^2)$ 的导数。令 $f(x)=2x+3$，$g(x)=3x^2$。则由此得 $f'(x)=2$，$g'(x)=6x$。按照（7.4），所求函数为

$$\frac{d}{dx}[(2x+3)(3x^2)] = (2x+3)(6x) + (3x^2)(2)$$

$$= 18x^2 + 18x$$

这个结果还可以通过首先将 $f(x)$ 与 $g(x)$ 相乘，然后再对乘积多项式求导加以检验。乘积多项式情况是 $f(x)g(x) = (2x+3)(3x^2)=6x^3+9x^2$，对其直接微分确实也得到同样的导数：$18x^2+18x$。 163

需要记住的重要之点是：两函数乘积的导数并不是两个孤立导数的简单乘积。因为这一结论与人们直观归纳所预期的结论不一致，所以我们对（7.4）进行证明。按照（6.13），当 $x \to N$ 时，$f(x)g(x)$ 的导数值应为

$$\frac{d}{dx}[f(x)g(x)]\big|_{x=N} = \lim_{x \to N}\frac{f(x)g(x) - f(N)g(N)}{x - N} \tag{7.5}$$

但,通过在分子中加和减 $f(x)g(N)$(从而原式大小不变),我们可将(7.5)右侧的差商变换如下:

$$\frac{f(x)g(x) - f(x)g(N) + f(x)g(N) - f(N)g(N)}{x - N}$$

$$= f(x)\frac{g(x) - g(N)}{x - N} + g(N)\frac{f(x) - f(N)}{x - N}$$

用此式代换(7.5)右边的差商并取极限,我们得到

$$\frac{d}{dx}[f(x)g(x)]\big|_{x=N} = \lim_{x \to N} f(x)\lim_{x \to N}\frac{g(x) - g(N)}{x - N} + \lim_{x \to N} g(N)\lim_{x \to N}\frac{f(x) - f(N)}{x - N} \tag{7.5'}$$

(7.5′)式中的四个极限表达式极易计算。第一项为 $f(N)$,第三项为 $g(N)$(常数的极限)。按(6.13),余下两项的极限分别为 $g'(N)$和 $f'(N)$。因而(7.5′)简化为

$$\frac{d}{dx}[f(x)g(x)]\big|_{x=N} = f(N)g'(N) + g(N)f'(N) \tag{7.5''}$$

并且,因 $N$ 表示 $x$ 的任意值,若我们以 $x$ 代替每个符号 $N$,(7.5″)依然成立。所以法则得证。

将此法则推广至三个函数的情况,我们有

$$\frac{d}{dx}[f(x)g(x)h(x)] = f'(x)g(x)h(x) + f(x)g'(x)h(x) + f(x)g(x)h'(x) \tag{7.6}$$

用文字描述即:三个函数乘积的导数等于第二个、第三个函数的积乘以第一个函数的导数,加上第一个、第三个函数的积乘以第二个

函数的导数，再加上第一个、第二个函数的积乘以第三个函数的导 164
数。这个结果可以通过反复应用(7.4)而推导出来。首先将积 $g(x)h(x)$ 当作单个函数，比如 $\phi(x)$，从而使原来三个函数的积变成两个函数的积，$f(x)\phi(x)$。这样，便可应用(7.4)了。求得 $f(x)\phi(x)$ 的导数后，我们可以对乘积 $g(x)h(x)\equiv\phi(x)$ 再应用(7.4)以得到 $\phi'(x)$。则可得到(7.6)。详细推导过程留给读者作练习。

这个法则的正确性是一回事，而其适用性则是另一回事。若我们可以将函数 $f(x)$ 和 $g(x)$ 相乘，然后再直接取积的导数，我们为什么还需要积的微分法则呢？对此问题的一个解释是这种直接取导数的步骤仅适用于特定的(数字或参数)函数，而如果函数是一般形式，则可应用积的微分法则。我们运用经济例子对此加以说明。

## 由平均收益函数求边际收益函数

若我们有下述形式的平均收益(AR)函数，

$$\mathrm{AR} = 15 - Q$$

首先用 $Q$ 乘以 AR 以便得到总收益($R$)函数，便可以求得边际收益(MR)函数：

$$R \equiv \mathrm{AR}\cdot Q = (15 - Q)Q = 15Q - Q^2$$

然后对 $R$ 取微分：

$$\mathrm{MR} \equiv \frac{dR}{dQ} = 15 - 2Q$$

但若 AR 函数以一般形式 $\mathrm{AR} = f(Q)$ 给出，则总收益函数也是一般形式的：

$$R = \mathrm{AR} \cdot Q = f(Q) \cdot Q$$

因而“相乘”的方法就完全失效了。然而，因 $R$ 是 $Q$ 的两个函数的积，即 $f(Q)$ 和 $Q$ 的积，积的微分法则便可以发生作用了。我们可以对 $R$ 微分以得到如下 MR 函数：

$$\mathrm{MR} \equiv \frac{dR}{dQ} = f(Q) \cdot 1 + Q \cdot f'(Q) = f(Q) + Qf'(Q) \tag{7.7}$$

但是，这个一般结果能使我们对 MR 有更多认识吗？答案是肯定的。回忆一下，$f(Q)$ 表示 AR 函数，我们将(7.7)重排并写成

$$\mathrm{MR} - \mathrm{AR} = \mathrm{MR} - f(Q) = Qf'(Q) \tag{7.7'}$$

这使我们了解 MR 与 AR 之间的重要关系：即它们之间总是相差 $Qf'(Q)$。

165　　我们继续考察 $Qf'(Q)$。其第一个成分 $Q$ 表示产出且总为非负。另一成分 $f(Q)$ 表示对 $Q$ 作出的 AR 曲线的斜率。因为“平均收益”与“价格”是同一事物的不同表述：

$$\mathrm{AR} \equiv \frac{R}{Q} \equiv \frac{PQ}{Q} \equiv P$$

因此，平均收益曲线也可以看作价格 $P$ 与产量 $Q$ 之间的关系曲线：$P = f(Q)$。因此，AR 曲线只是对该厂商产品需求曲线的反向曲线，$P$ 和 $Q$ 轴对换后所绘成的需求曲线。在完全竞争条件下，AR 曲线为水平直线，从而 $f'(Q) = 0$，由(7.7′)，对所有可能的 $Q$ 值，$\mathrm{MR} - \mathrm{AR} = 0$。因此，MR 曲线与 AR 曲线必定是重合的。另一方面，在不完全竞争条件下，AR 曲线一般为向下倾斜的曲线，如图 7.2 所示，从而 $f'(Q) < 0$，且由(7.7′)，对所有正的产出水平，$\mathrm{MR} - \mathrm{AR} < 0$。在这种情况下，MR 曲线必位于 AR 曲线之下。

刚才阐述的结论实质上是定性的。它仅涉及两个曲线的相对位置。但(7.7′)也提供了下述定量的信息:在任何产出水平 $Q$,

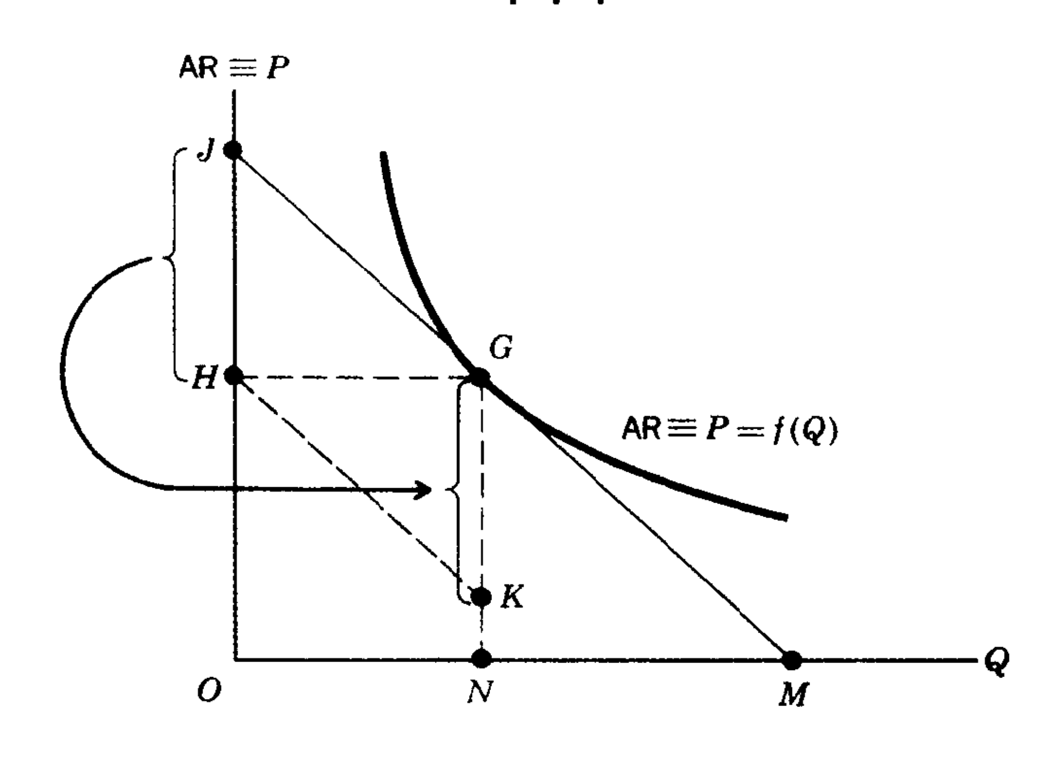



图 7.2

MR 曲线总是精确地低于 AR 曲线 $Qf'(Q)$ 的距离。我们再看一下图 7.2,并考察特定的产出水平 $N$。在该产出水平,表达式 $Qf'(Q)$ 则变成 $Nf'(N)$。若我们能求出图中 $Nf'(N)$ 的大小,我们便可以知道平均收益点 $G$ 与相应的边际收益点之间的距离。

$N$ 的大小已经给定。$f'(N)$ 仅为 AR 曲线在点 $G$(其中 $Q = N$)的斜率,即切线 $JM$ 的斜率,它是距离 $OJ$ 与 $OM$ 的比率,即 $OJ/OM$。但我们看到 $OJ/OM = HJ/HG$,此外,$HG$ 正好等于所考察的产出水平 $N$ 的数量。因此,距离 $Nf'(N)$,即在产出水平 $N$ 时 MR 曲线位于 AR 曲线之下的距离,为 166

$$Nf'(N) = HG\frac{HJ}{HG} = HJ$$

据此,若我们直接在点 $G$ 下标出垂线距离 $KG = HJ$,则点 $K$ 必定

是 MR 曲线上的点。(准确绘出 $KG$ 的简单方法是画一条通过点 $H$ 并与 $JG$ 平行的直线;点 $K$ 为该线与垂线 $NG$ 的交点。)

可用同样的方法确定 MR 曲线上其它的点。对于 AR 曲线上任意选定的点 $G'$,我们所要做的是,首先过 $G'$ 画一条 AR 曲线的切线,它与纵轴交于 $J'$。然后过 $G'$ 向纵轴画一条水平线,它与纵轴交于 $H'$。若我们直接在 $G'$ 点下方标出垂直距离 $K'G' = H'J'$,则 $K'$ 将位于 MR 曲线上。这就是由已知的 AR 曲线导出 MR 曲线的作图方法。严格地讲,精确地绘出切线需要知道在相应产出水平的导数值,即 $f'(N)$。所以刚才描述的图形法是不能独立存在的。一个主要的例外是线性平均收益的情况,AR 曲线上任何一点的切线都是 AR 曲线自己,所以没有必要绘出任何切线。所以上述图形法可以直接应用。

## 商的积分法则

两个函数的商 $f(x)/g(x)$ 的导数为

$$\frac{d}{dx}\frac{f(x)}{g(x)} = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$$

在右侧表达式的分子中,我们看到两个乘积项,每项中仅涉及两个原函数中的一个函数的导数。注意:$f'(x)$ 出现在正的一项中,$g'(x)$ 出现在负的一项中。分母由函数 $g(x)$ 的平方构成,即 $g^2(x) \equiv [g(x)]^2$。

**例 6** $$\frac{d}{dx}\left(\frac{2x-3}{x+1}\right) = \frac{2(x+1)-(2x-3)(1)}{(x+1)^2} = \frac{5}{(x+1)^2}$$

**例 7** $$\frac{d}{dx}\left(\frac{5x}{x^2+1}\right) = \frac{5(x^2+1)-5x(2x)}{(x^2+1)^2} = \frac{5(1-x^2)}{(x^2+1)^2}$$

**例 8** $$\frac{d}{dx}\left(\frac{ax^2+b}{cx}\right) = \frac{2ax(cx)-(ax^2+b)(c)}{(cx)^2}$$

$$= \frac{c(ax^2 - b)}{(cx)^2} = \frac{ax^2 - b}{cx^2}$$

这一法则可证明如下。对任意 $x = N$,我们有 167

$$\left. \frac{d}{dx} \frac{f(x)}{g(x)} \right|_{x=N} = \lim_{x \to N} \frac{f(x)/g(x) - f(N)/g(N)}{x - N} \tag{7.8}$$

极限符号后面的商表达式可重写成如下形式

$$\frac{f(x)g(N) - f(N)g(x)}{g(x)g(N)} \frac{1}{x - N}$$

在分子中同时加、减 $f(N)g(N)$,并重排,我们可以将其变换为

$$\frac{1}{g(x)g(N)}\left[\frac{f(x)g(N) - f(N)g(N) + f(N)g(N) - f(N)g(x)}{x - N}\right]$$

$$= \frac{1}{g(x)g(N)}\left[g(N)\frac{f(x) - f(N)}{x - N} - f(N)\frac{g(x) - g(N)}{x - N}\right]$$

将此结果代入(7.8)并取极限,我们有

$$\left. \frac{d}{dx} \frac{f(x)}{g(x)} \right|_{x=N} = \lim_{x \to N} \frac{1}{g(x)g(N)}\left[\lim_{x \to N} g(N) \lim_{x \to N} \frac{f(x) - f(N)}{x - N} - \lim_{x \to N} f(N) \lim_{x \to N} \frac{g(x) - g(N)}{x - N}\right]$$

$$= \frac{1}{g^2(N)}[g(N)f'(N) - f(N)g'(N)]$$

[由(6.13)]

通过用 $x$ 代替符号 $N$。可使上式一般化,因为 $N$ 表示任意 $x$ 值。商的微分法则得证。

## 边际成本与平均成本的关系

作为商的微分法则在经济中的一个应用,我们考察产出变化时平均成本的变化率。

给定总成本函数 $C = C(Q)$,平均成本($AC$)函数为 $Q$ 的两个

函数的商，因为 $AC = C(Q)/Q$，只要 $Q>0$ 便有定义。因此，通过对 $AC$ 求导可求出 $AC$ 对 $Q$ 的变化率：

$$\frac{d}{dQ}\frac{C(Q)}{Q}=\frac{[C'(Q)\cdot Q-C(Q)\cdot 1]}{Q^2}=\frac{1}{Q}\left[C'(Q)-\frac{C(Q)}{Q}\right] \tag{7.9}$$

由此可得，对于 $Q>0$，

$$\frac{d}{dQ}\frac{C(Q)}{Q} \gtreqless 0 \quad \text{当且仅当 } C'(Q) \gtreqless \frac{C(Q)}{Q} \tag{7.10}$$

168

由于导数 $C'(Q)$ 表示边际成本(MC)函数，$C(Q)/Q$ 表示平均成本(AC)函数，(7.10)的经济含义为：当且仅当边际成本曲线位于AC曲线的上方、与之相交，或者位于其下方时，AC曲线的斜率将为正、零，或者为负。这个结果在图7.3中加以说明。其中MC和AC函数是按下面特定的总成本函数绘出的

$$C = Q^3 - 12Q^2 + 60Q$$

在 $Q=6$ 的左侧，AC是递减的，MC位于AC的下方。在该点右侧，情况恰好相反。而在 $Q=6$，AC的斜率为零，MC和AC的值相等。①

(7.10)中的定性结论是完全按成本函数来表述的。但若我们将 $C(Q)$ 看作其它任何可微的总函数，把 $C(Q)/Q$ 和 $C'(Q)$ 看作相应的平均和边际函数，上述结论依然是有效的。所以，这个结果揭示了一般的边际——平均关系。我们特别应指出的是：在与图7.2相联系的讨论中，当AR向下倾斜时MR位于AR下方，仅是

① 注意，(7.10)并不表明当AC斜率为负时，MC的斜率也必然为负；它只表明此时AC必定大于MC。在图7.3中，如 $Q=5$ 时，AC递减，而MC增加，所以MC斜率为正。

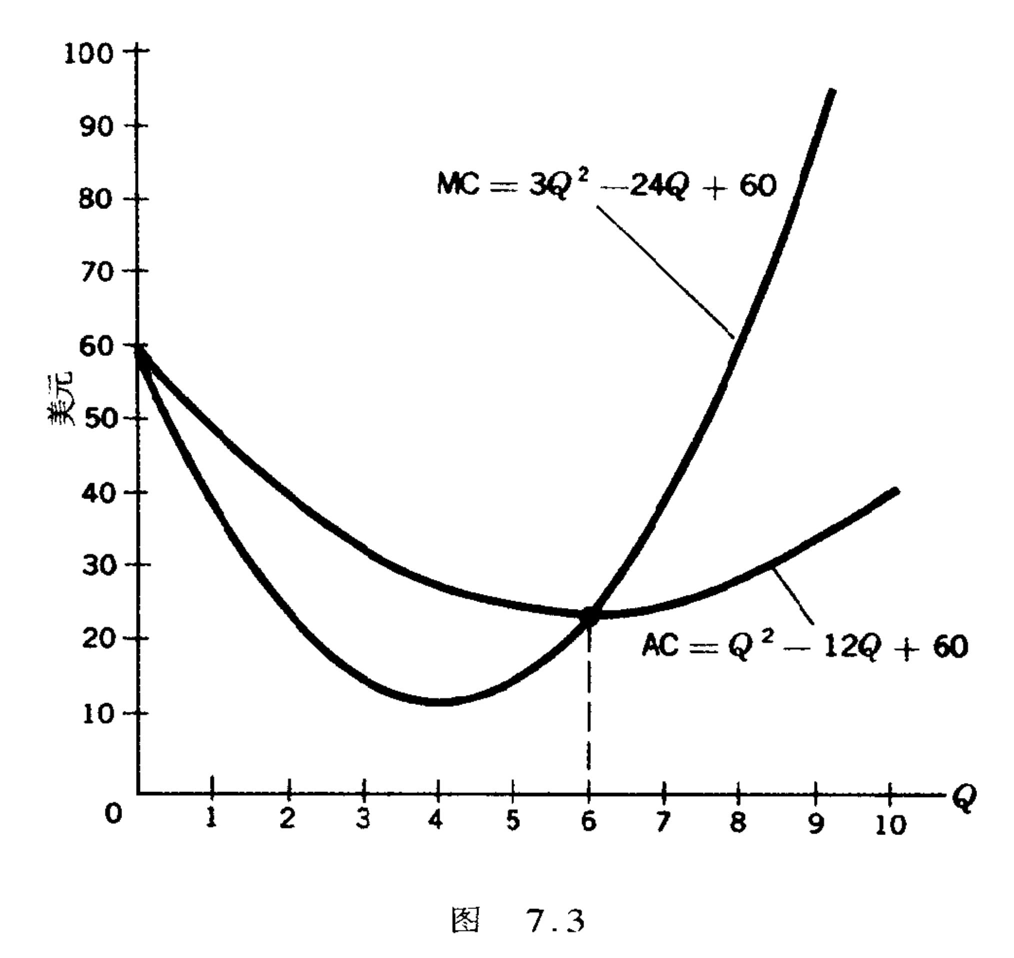



图 7.3

(7.10)一般性结论中的一个特例。

## 练习 7.2



---

1 给定总成本函数 $C=Q^3-5Q^2+14Q+75$,写成可变成本(VC)函数,求 VC 函数的导数,并解释其经济含义。

2 已知平均成本函数 $AC=Q^2-4Q+214$,求 MC 函数。这个已知函数更适合于作长期成本函数或短期成本函数吗?为什么?

3 运用积的微分法则求下列导数

| | |
|---|---|
| ($a$) $(9x^2-2)(3x+1)$ | ($d$)$(ax-b)(cx^2)$ |
| ($b$) $(3x+11)(6x^2-5x)$ | ($e$)$(2-3x)(1+x)(x+2)$ |
| ($c$) $x^2(4x+6)$ | ($f$)$(x^2+3)x^{-1}$ |

4 ($a$)给定 AR$=60-3Q$,绘出平均收益曲线,并运用图 7.2 所用之方法求出 MR 曲线。

($b$) 用数学方法由给定的 AR 函数求总收益函数和边际收益函数。

($c$) ($a$) 中以图形方法推导出的 MR 曲线与($b$)中以数学方法推导出的 MR 曲线一致吗?

($d$) 比较 AR 和 MR 函数,关于其相对斜率可得出何结论?

5 给出下述一般结论的数学证明:给定线性平均曲线,其相应的边际曲线必定与平均曲线有相同的截距,而其陡度是平均曲线的二倍。

6 用下述方法证明(7.6)的结论:先将 $g(x)h(x)$ 看作单个函数,即令 $g(x)h(x)\equiv\phi(x)$,然后再应用(7.4)的积的微分法则。

7 求下列函数的导数

($a$) $(x^2+3)/x$　　($c$) $4x/(x+5)$

($b$) $(x+7)/x$　　($d$) $(ax^2+b)/(cx+d)$

8 已知函数 $f(x)=ax+b$,求下列函数的导数:

($a$) $f(x)$　($b$) $xf(x)$　($c$) $1/f(x)$　($d$) $f(x)/x$

---

## 7.3 包含不同自变量的函数的微分法则

上一节,我们讨论了具有同一变量的两个或两个以上可微函数的和、差、积、商的微分法则。现在我们考察两个或两个以上可微函数,每个函数具有不同自变量的情况。

### 链微分法则

若我们有函数 $z=f(y)$,其中 $y$ 又是另一变量 $x$ 的函数,比如 $y=g(x)$,则 $z$ 对 $x$ 的导数等于 $z$ 对 $y$ 的导数乘以 $y$ 对 $x$ 的导数,170用符号表示如下:

$$\frac{dz}{dx}=\frac{dz}{dy}\frac{dy}{dx}=f'(y)g'(x) \tag{7.11}$$

这一微分法则称作链微分法则，这个名称是很直观的。给定 $\Delta x$，必定通过函数 $y=g(x)$ 产生相应的 $\Delta y$；而 $\Delta y$ 必定通过函数 $z=f(y)$ 产生 $\Delta z$。因而存在如下“链式反应”：

$$\Delta x \xrightarrow{\text{通过 } g} \Delta y \xrightarrow{\text{通过 } f} \Delta z$$

在这个链的两个衔接处，产生两个差商 $\Delta y/\Delta x$ 和 $\Delta z/\Delta y$，但它们相乘时，$\Delta y$ 本身会消去，从而我们得到

$$\frac{\Delta z}{\Delta y}\frac{\Delta y}{\Delta x}=\frac{\Delta z}{\Delta x}$$

一个将 $\Delta z$ 与 $\Delta x$ 联系起来的差商。若我们取当 $\Delta x \to 0$ 时这些差商的极限，每一个差商都将变成导数，即我们有 $(dz/dy)(dy/dx)=dz/dx$。这与(7.11)中的结果完全一致。

考虑到函数 $y=g(x)$，我们可以将函数 $z=f(y)$ 表示成 $z=f[g(x)]$。其中两个函数符号 $f$ 和 $g$ 相连出现，表明它是一个复合函数(函数的函数)。由于这个原因，链微分法则也称复合函数微分法则或复函数微分法则。

链微分法则可直接推广至三个或更多函数中去。若我们有 $z=f(y)$，$y=g(x)$，$x=h(w)$，则

$$\frac{dz}{dw}=\frac{dz}{dy}\frac{dy}{dx}\frac{dx}{dw}=f'(y)g'(x)h'(w)$$

类似的推广也适用于更多函数的情况。

**例 1** 若 $z=3y^2$，其中 $y=2x+5$，则

$$\frac{dz}{dx}=\frac{dz}{dy}\frac{dy}{dx}=6y(2)=12y=12(2x+5)$$

**例 2** 若 $z=y-3$，其中 $y=x^3$，则

$$\frac{dz}{dx}=1(3x^2)=3x^2$$

**例3** 当必须对像 $z=(x^2+3x-2)^{17}$ 这样的函数求导时，我们便可以更好地理解链微分法则的价值。如果我们不掌握此法则，$dz/dx$ 只能通过先将表达式的 17 次幂展开后，才能极其繁琐地求得。但运用链微分法则，我们便可以走捷径：定义一个新的中间变量 $y=x^2+3x-2$，从而得到联结成链的两个函数：

171

$$z=y^{17} \quad 和 \quad y=x^2+3x-2$$

则导数 $dz/dx$ 可求出如下：

$$\frac{dz}{dx}=\frac{dz}{dy}\frac{dy}{dx}=17y^{16}(2x+3)=17(x^2+3x-2)^{16}(2x+3)$$

**例4** 给定某厂商的总收益函数 $R=f(Q)$，其中产出 $Q$ 为劳动投入 $L$ 的函数，或 $Q=g(L)$。由链微分法则，我们有

$$\frac{dR}{dL}=\frac{dR}{dQ}\frac{dQ}{dL}=f'(Q)g'(L)$$

将其转换成经济术语，$dR/dQ$ 是 MR 函数，$dQ/dL$ 为劳动的边际物质产品（$MPP_L$）函数。类似地，$dR/dL$ 为劳动的边际收益产品（$MRP_L$）函数。因此，上述结果给出了经济学中著名结论的数学表述：$MRP_L=MR\cdot MPP_L$。

## 反函数微分法则

若函数是一个一对一映射的函数，即若函数使得不同的 $x$ 值总对应不同的 $y$ 值，则函数 $f$ 具有反函数 $x=f^{-1}(y)$（读作 $x$ 为 $y$ 的反函数）。这里，符号 $f^{-1}$ 像导函数符号 $f'$ 一样，是一个函数符号，表示一个与函数 $f$ 相联系的一个函数，它并非函数 $f(x)$ 的倒数。

反函数的存在实际上意味着，在此情况下，不仅给定的 $x$ 值产生唯一的 $y$ 值，即 $y=f(x)$，而且，给定的 $y$ 值也产生一个惟一的 $x$ 值。我们给出一个非数字的例子：在一个一夫一妻制社会中，我们

可以把所有丈夫的集合对所有妻子的集合看作一个一对一的映射。每个丈夫具有一个妻子，每个妻子也仅有唯一的丈夫。相反，所有父亲的集合对所有儿子的集合则不是一对一的映射，因为一个父亲可能拥有多个儿子，而每个儿子则仅有唯一的父亲。

当 $x$ 和 $y$ 指具体的数时，一对一映射的性质被看作是单调函数这类函数所特有的性质。给定函数 $f(x)$，若自变量 $x$ 的值连续增大时，$f(x)$ 的值也总是连续地增大，即若

$$x_1 > x_2 \Rightarrow f(x_1) > f(x_2)$$

则函数 $f$ 称作增(或单调递增)函数。[1] 若 $x$ 的连续增加总是导致 172 $f(x)$ 的连续递减，即若

$$x_1 > x_2 \Rightarrow f(x_1) < f(x_2)$$

则该函数被称作减(或单调递减)函数。在上述每种情况下，反函数 $f^{-1}$ 存在。

确定给定函数 $y = f(x)$ 单调性的一种实用的方法是检验导数 $f'(x)$ 对所有的 $x$ 值，是否总是具有同样的代数符号(不能为零)。从几何上看，这意味着曲线或者总是向上倾斜，或者总是向下倾斜。因此，一个总是具有负的斜率的厂商要求曲线 $Q = f(P)$ 是单调递减的。所以，它具有反函数 $P = f^{-1}(Q)$。如前面所提到的那样，它给出了厂商的平均收益曲线，因为 $P \equiv \mathrm{AR}$。

---

① 有些作者更喜欢将增函数定义为具有下述性质的函数：

$$x_1 > x_2 \Rightarrow f(x_1) \geqslant f(x_2) \qquad [\text{弱不等式}]$$

而用严格增函数描述这种情况

$$x_1 > x_2 \Rightarrow f(x_1) > f(x_2) \qquad [\text{严格不等式}]$$

在这种用法中，上升的阶段函数也看作增函数(尽管不是严格递增的)，尽管事实上它是水平线段。我们在本书中不使用这种用法。相反，我们不将上升的阶段函数视为增函数，而视为非减函数。同样，我们不将下降的阶段函数视为减函数，而视为非增函数。

**例5** 函数

$$y = 5x + 25$$

具有导数 $dy/dx = 5$,无论 $x$ 取何值,它始终为正。因此,此函数也是单调函数。(它是单调递增函数,因导数为正。)所以,反函数存在。在本例中,解已知方程 $y = 5x + 25$,求出 $x$,便很容易求出了反函数。反函数为

$$x = \frac{1}{5}y - 5$$

令人感兴趣的是,这个反函数也是单调递增的,因为对所有的 $y$ 值,$dx/dy = 1/5 > 0$。

一般而言,若反函数存在,原函数和反函数必定都是单调的。进而,若 $f^{-1}$为函数 $f$ 的反函数,则 $f$ 必为 $f^{-1}$的反函数。即 $f$ 和 $f^{-1}$必然互为反函数。

很容易验证,$y = f(x)$的图形与 $x = f^{-1}(y)$的图形是同一图形,但两个坐标轴恰好颠倒了。若我们将 $f^{-1}$图形的 $x$ 轴置于 $f$ 图形的 $x$ 轴上,(对 $y$ 轴也作类似处理),则两条曲线将重合。若 $f^{-1}$
173
图形中的 $x$ 轴置于 $f$ 图形中的 $y$ 轴上(而反之亦然)则两条曲线以通过原点的 45°直线为轴,呈镜像对称。一旦原函数 $f$ 的图形给定,镜像关系为我们提供了反函数 $f^{-1}$图形的简单方法。(读者可用例 5 中的两个函数尝试作图。)

反函数的微分法则为

$$\frac{dx}{dy} = \frac{1}{dy/dx}$$

这意味着反函数的导数是原函数导数的倒数。因此,$dx/dy$ 一定与 $dy/dx$ 具有相同的正负号。从而若 $f$ 是增(减)函数,则 $f^{-1}$必然也为增(减)函数。

要验证此微分法则,我们可以回到例5,其中已求得 $dy/dx=5$, $dx/dy=1/5$。两个导数确实互为倒数,并且具有相同的符号。

在这个简单的例子中,反函数相对易于求得,从而导数 $dx/dy$ 可直接从反函数求出。但正如下一个例子将要显示的那样,反函数有时是很难直接表示清楚的。所以直接微分是行不通的。因此,反函数微分法则的价值变得更为明显。

**例6** 给定 $y=x^5+x$,求 $dx/dy$。首先,因为对任意 $x$ 值,

$$\frac{dy}{dx}=5x^4+1>0$$

所以这个函数是单调递增的,且反函数存在。解此方程求出 $x$ 并不是一个轻松的任务,但通过运用反函数微分法则,便可很快求出反函数的导数:

$$\frac{dx}{dy}=\frac{1}{dy/dx}=\frac{1}{5x^4+1}$$

严格地讲,仅当函数包含一对一的映射时,才可应用反函数的微分法则。但事实上,我们确实还有一些选择余地。例如,当涉及U形曲线(非单调)时,我们可以将曲线向下和向上倾斜的部分分别看成两个单独的函数,每个函数具有限定的定义域,且在限定定义域中是单调的,所以便可对其应用反函数的微分法则了。

## 练习 7.3

1 已知 $y=u^3+1$,其中 $u=5-x^2$,运用链微分法则求 $dy/dx$。

2 已知 $w=ay^2$,$y=bx^2+cx$,运用链微分法则求 $dw/dx$。 174

3 运用链微分法则求下列函数的导数:

(a) $y=(3x^2-13)^3$　(b) $y=(8x^3-5)^9$　(c) $y=(ax+b)^4$

4　已知 $y=(16x+3)^{-2}$，运用链微分法则求 $dy/dx$。然后将函数改写成 $y=1/(16x+3)^2$，并用商的微分法则求 $dy/dx$。两个答案一致吗？

5　给定 $y=7x+21$，求其反函数。然后求 $dy/dx$ 和 $dx/dy$，并验证反函数微分法则。然后验证两函数图形彼此间存在镜像关系。

6　下列函数是单调函数吗？

(a) $y=-x^6+5\ (x>0)$　(b) $y=4x^5+x^3+3x$ 对于每个单调函数，用反函数微分法则求 $dx/dy$。

---

## 7.4　偏微分

迄今为止，我们仅考察了具有单个自变量的函数的导数。但在比较静态分析中，我们可能遇到模型中出现几个参数的情况，在此情况下内生变量的均衡值可能不止是一个参数的函数。因此，作为将导数概念用于比较静态分析的最后准备，我们必须学习如何求多于一个变量的函数的导数问题。

### 偏导数

我们考察函数

$$y=f(x_1,x_2,\cdots,x_n) \tag{7.12}$$

其中变量 $x_i(i=1,2,\cdots,n)$ 彼此完全独立，所以，每个变量自己变化，不会对其它变量产生影响。若变量 $x_1$ 发生变化 $\Delta x_1$，其它变量 $x_2,\cdots,x_n$ 保持不变，则 $y$ 会产生相应的变化 $\Delta y$。此情况下的差商可以表示为

$$\frac{\Delta y}{\Delta x_1}\ \frac{f(x_1+\Delta x_1,x_2,\cdots,x_n)-f(x_1,x_2,\cdots,x_n)}{\Delta x_1} \tag{7.13}$$

若我们取当 $\Delta x_1 \to 0$ 时 $\Delta y/\Delta x_1$ 的极限,极限将构成一个导数。我们称其为 $y$ 对 $x_1$ 的偏导数,以表示当取这个特定导数时,函数中所有其它自变量均保持不变。当其它自变量有无穷小变化时,还可以定义类似的偏导数。取偏导数的过程称作偏微分。

偏导数可用不同的符号表示。这里我们不用符号 $d$(像 $dy/dx$),而用希腊字母 $\delta$(小写的 *delta*)的变体 $\partial$ 表示偏导数。所以我们现在写成 $\partial y/\partial x_i$,读作"$y$ 对 $x_i$ 的偏导数"。偏导数符号有时也被写成$\frac{\partial}{\partial x_i}y$;在此情况下,$\partial/\partial x_i$ 部分可以视作一个算符,它告诉我们取某一函数对 $x_i$ 的偏导数。因为这里所涉及的函数在(7.12)中以 $f$ 表示,所以将此偏导数写成 $\partial f/\partial x_i$ 也是可以的。 175

我们前面曾用符号 $f'(x)$表示导数,偏导数也有对应的简写表示吗?答案是肯定的。但这里用 $f_1, f_2$ 等等,而不用 $f'$,其中下标标明允许变化的自变量。若函数(7.12)以未带下标的自变量表示,如 $y = f(u, v, w)$,则偏导数以 $fu, fv, fw$ 表示,而不用 $f_1, f_2, f_3$ 表示。

与上面这些记号一致,根据(7.12)和(7.13),我们现在可以将下式

$$f_1 \equiv \frac{\partial y}{\partial x_1} \equiv \lim_{\Delta x_1 \to 0} \frac{\Delta y}{\Delta x_1}$$

定义为函数 $fn$ 个偏导数集合中的第一个偏导数。

## 求偏导数的技巧

偏导数与前面讨论的微分的基本差别在于,我们必须保持$(n-1)$个变量不变,而只允许一个变量变化。由于我们已掌握了在微分中如何处理常数的方法,所以,求偏微分不应存在什么问题。

**例 1**　已知 $y=f(x_1,x_2)=3x_1^2+x_1x_2+4x_2^2$,求偏导数。我们必须记住,在求 $\partial y/\partial x_1$(或 $f_1$)时,应把 $x_2$ 视为常数。因此,若 $x_2$ 是一个加和常数,在求导时将被去掉;但若它是一个乘积常数(如 $x_1x_2$ 项中的 $x_2$),它仍将被保留。因此,我们有

$$\frac{\partial y}{\partial x_1}\equiv f_1=6x_1+x_2$$

类似地,把 $x_1$ 作为常数,我们求得

$$\frac{\partial y}{\partial x_2}\equiv f_2=x_1+8x_2$$

注意,像原函数 $f$ 一样,两个偏导数均为变量 $x_1$ 和 $x_2$ 的函数。即我们可以将其写成两个导函数

$$f_1=f_1(x_1,x_2) \quad 和 \quad f_2=f_2(x_1,x_2)$$

例如,在函数 $f$ 定义域中的点 $(x_1,x_2)=(1,3)$,偏导数将取下面两个具体值:

$$f_1(1,3)=6(1)+3=9 \quad 和 \quad f_2(1,3)=1+8(3)=25$$

176

**例 2**　已知 $y=f(u,v)=(u+4)(3u+2v)$,可用积的微分法则求偏导数。保持 $v$ 不变,我们有

$$f_u=(u+4)(3)+1(3u+2v)=2(3u+v+6)$$

类似地,保持 $u$ 不变,我们求得

$$f_v=(u+4)(2)+0(3u+2v)=2(u+4)$$

当 $u=2,v=1$ 时,这两个偏导数取值如下:

$$f_u(2,1)=2(13)=26 \quad 和 \quad f_v(2,1)=2(6)=12$$

**例 3**　给定 $y=(3u-2v)/(u^2+3v)$,可用商的微分法则求偏导数

$$\frac{\partial y}{\partial u}=\frac{3(u^2+3v)-2u(3u-2v)}{(u^2+3v)^2}=\frac{-3u^2+4uv+9v}{(u^2+3v)^2}$$

$$\frac{\partial y}{\partial v}=\frac{-2(u^2+3v)-3(3u-2v)}{(u^2+3v)^2}=\frac{-u(2u+9)}{(u^2+3v)^2}$$

## 偏导数的几何解释

偏导数作为一种特殊类型的导数,是某个变量瞬间变化率的度量,在几何上它也对应于特定曲线的斜度。

我们考察生产函数 $Q=Q(K,L)$,其中 $Q$、$K$、$L$ 分别表示产出、资本投入和劳动投入。这个函数是(7.12)的两变量形式,即 $n=2$。因此我们可以定义两个偏导数 $\partial Q/\partial K$(或 $Q_K$)和 $\partial Q/\partial L$(或 $Q_L$)。偏导数 $Q_K$ 与资本有无穷小变化时的产出变化率相联系,而劳动投入保持不变。所以,$Q_K$ 表示资本的边际物质产品($MPP_K$)函数。类似地,偏导数 $Q_L$ 是 $MPP_L$ 函数的数学表示。

从几何上看,如图 7.4 所示,生产函数 $Q=Q(K,L)$ 可用三维空间中的一个**生产曲面**来描绘。变量 $Q$ 绘成纵轴,所以对底平面($KL$ 平面)上任意一点$(K,L)$,曲面的高度将标明产出 $Q$。函数的定义域由底平面的整个非负象限构成,但基于我们的分析目的,考察此非负象限中的一个子集——矩形 $OK_0BL_0$ 已经足够了。因此,在图 7.4 中仅显示生产曲线中很小一部分。

现在我们将资本维持在水平 $K_0$,仅考察投入 $L$ 的变化。令 $K=K_0$,除了线段 $K_0B$ 上的点以外,(缩小了的)定义域中所有的点均与所考察之问题无关。同样,只有曲线 $K_0CDA$(生产曲面的截线)与我们现在所讨论的问题有关。此曲线表示对固定资本数量 $K=K_0$ 时的劳动的总物质产品($TPP_L$)曲线。因此,我们可由其斜率读出当 $K$ 不变时,产出对 $L$ 变化的变化率。很明显,所以像 $K_0CDA$ 这样的曲线是偏导数 $Q_L$ 在几何上的对应物。我们再

177

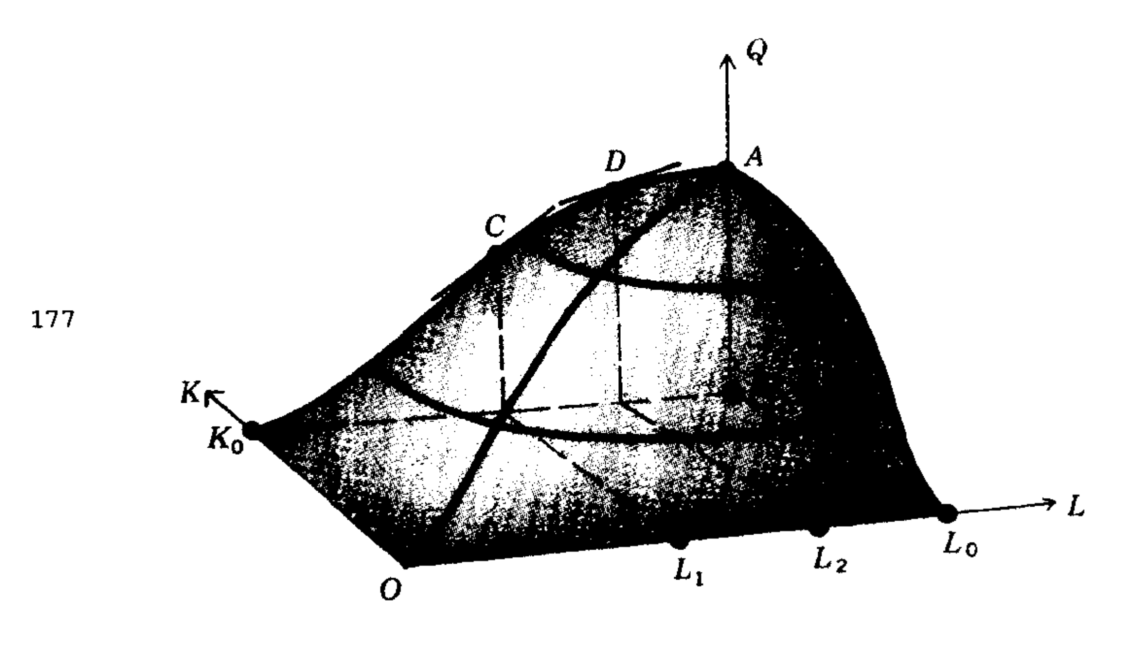



图 7.4

一次强调,劳动的总物质产品($TPP_L$)曲线对应于劳动的边际物质产品($MPP_L = Q_L$)曲线。

前面已提到,偏导数是原函数中所有自变量的函数,$Q_L$ 作为 $L$ 的函数从 $K_0CDA$ 曲线可立即看出,当 $L = L_1$ 时,$Q_L$ 的值等于曲线在点 $C$ 的斜率;但当 $L = L_2$ 时,$Q_L$ 的值则等于在点 $D$ 的斜率。那么,为什么 $Q_L$ 也是 $K$ 的函数呢? 答案是,$K$ 可以固定在不同水平,对于不同的固定的 $K$ 水平,将会产生不同的 $TPP_L$ 曲线(生产曲线的不同的截线),从而必然会对导数 $Q_L$ 产生影响。因此,$Q_L$ 也是 $K$ 的函数。

类似的解释也适用于偏导数 $Q_K$。若劳动投入固定不变(比如固定在水平 $L_0$),则 $L_0B$ 为定义域中的相关子集;曲线 $L_0A$ 表示生产曲面上的相关子集。偏导数 $Q_K$ 则可解释为曲线 $L_0A$ 的斜率,要注意在图 7.4 中,$K$ 轴由东南向西北方向延伸。还应指出的是,$Q_K$ 也是变量 $L$ 和 $K$ 的函数。

## 练习 7.4

1 求下列每个函数的 $\partial y/\partial x_1, \partial y/\partial x_2$：

(a) $y = 2x_1^3 - 11x_1^2x_2 + 3x_2^2$　　(c) $y = (2x_1 + 3)(x_2 - 2)$

(b) $y = 7x_1 + 5x_1x_2^2 - 9x_2^3$　　(d) $y = (4x_1 + 3)/(x_2 - 2)$

178

2 求下列函数的 $f_x$ 和 $f_y$：

(a) $f(x,y) = x^2 + 5xy - y^3$　　(c) $f(x,y) = \dfrac{2x - 3y}{x + y}$

(b) $f(x,y) = (x^2 - 3y)(x - 2)$　　(d) $f(x,y) = \dfrac{x^2 - 1}{xy}$

3 由上述问题的答案，求每一函数的 $f_x(1,2)$，即求当 $x = 1, y = 2$ 时，偏导数 $f_x$ 的值。

4 给定生产函数 $Q = 96K^{0.3}L^{0.7}$，求 $MPP_K$ 函数和 $MPP_L$ 函数。$MPP_K$ 仅是 $K$ 的函数，还是既是 $K$ 的函数，又是 $L$ 的函数？$MPP_L$ 呢？

5 若某人的效用函数取下述形式：

$$U = U(x_1, x_2) = (x_1 + 2)^2(x_2 + 3)^3$$

其中 $U$ 为总效用函数，$x_1$、$x_2$ 为所消费商品的数量，

(a) 求每一商品的边际效用函数。

(b) 求当消费的每个商品均为 3 个单位时，第一个商品的边际效用值。

# 7.5 微分在比较静态分析中应用

掌握了各种微分法则的知识后，我们终于可以解决比较静态分析所提出的问题：即当任意外生变量或参数发生变化时，内生变量的均衡值将如何变化。

### 市场模型

首先，我们再一次考察(3.1)式简单的单一商品市场模型。此

模型可以写成下述形式的两个方程：

$$Q = a - bP \qquad (a, b > 0) \qquad [需求]$$

$$Q = -c + dP \qquad (c, d > 0) \qquad [供给]$$

其解为

$$\bar{P} = \frac{a + c}{b + d} \tag{7.14}$$

$$\bar{Q} = \frac{ad - bc}{b + d} \tag{7.15}$$

这两个解可以看作简化形式：两个内生变量被简化为四个独立参数 $a$、$b$、$c$、$d$ 的显示表示。

为了解某一参数的一个无穷小变化如何影响 $\bar{P}$ 值，我们仅需对(7.14)中每一参数求偏导数。若偏导数的正负号，比如 $\partial\bar{P}/\partial a$ 179 的正负号可由关于参数的给定信息来确定，我们便会知道当参数 $a$ 变化时 $\bar{P}$ 的变化方向。这个分析给出了定性的结论。若 $\partial\bar{P}/\partial a$ 的大小可以确定，则此分析给出了定量的结论。

类似地，像 $\partial\bar{Q}/\partial a$ 一样，由 $\bar{Q}$ 对每一参数的偏导数，我们还可以得出定性或定量的结论。然而，为了避免误解，我们应在导数 $\partial\bar{Q}/\partial a$ 和 $\partial Q/\partial a$ 之间作明确区分。后一导数是一个仅适于需求函数的概念，不考虑供给函数的影响。而导数 $\partial\bar{Q}/\partial a$ 由(7.15)的均衡数量求得，(7.15)本质上是模型的解，将供给和需求的相互作用综合到了一起。为强调这一区别，我们将 $\bar{P}$ 和 $\bar{Q}$ 对参数的导数称作**比较静态导数**。

现在集中讨论 $\bar{P}$，我们由(7.14)可得出四个偏导数：

$$\frac{\partial\overline{P}}{\partial a}=\frac{1}{b+d}\quad\left[\text{参数 } a \text{ 的系数是}\frac{1}{b+d}\right]$$

$$\frac{\partial\overline{P}}{\partial b}=\frac{0(b+d)-1(a+c)}{(b+d)^2}=\frac{-(a+c)}{(b+d)^2}\quad[\text{商的微分法则}]$$

$$\frac{\partial\overline{P}}{\partial c}=\frac{1}{b+d}\left(=\frac{\partial\overline{P}}{\partial a}\right)$$

$$\frac{\partial\overline{P}}{\partial d}=\frac{0(b+d)-1(a+c)}{(b+d)^2}=\frac{-(a+c)}{(b+d)^2}\left(=\frac{\partial\overline{P}}{\partial b}\right)$$

在此模型中,因所有参数被限定为正,所以我们可以得出结论

$$\frac{\partial\overline{P}}{\partial a}=\frac{\partial\overline{P}}{\partial c}>0\quad\text{和}\quad\frac{\partial\overline{P}}{\partial b}=\frac{\partial\overline{P}}{\partial d}<0\tag{7.16}$$

为充分理解(7.16)中的结论,我们观察图 7.5,其中每个图表示一个参数的变化。同前面一样,我们把 $Q$(而非 $P$)画作纵轴。

图 7.5$a$ 给出了参数 $a$ 增至 $a'$ 时的图形表示,这意味着需求曲线有更高的纵截距,由于参数 $b$(斜率参数)未变,$a$ 的增加导致需求曲线由 $D$ 平移至 $D'$。$D'$ 与供给曲线 $S$ 的交点决定了均衡价格 $\overline{P}'$,它高于原均衡价格 $\overline{P}$。这便证明了 $\partial\overline{P}/\partial a>0$ 的结论,尽管为清晰起见,我们在图 7.5$a$ 中所绘出的参数 $a$ 的变化比导数概念所指的变化要大。

图 7.5$c$ 与图 7.5$a$ 有类似的解释。但因发生变化的是参数 $c$,所以结果是供给曲线的平行移动。注意,这里是向下的平行移 180 动,因为供给曲线具有纵截距 $-c$,因此,$c$ 的增加会使纵截距比如从 $-2$ 变为 $-4$。这个图形比较静态分析的结果是 $\overline{P}'$ 大于 $\overline{P}$,与导数 $\partial\overline{P}/\partial c$ 具有正号是一致的,这也是我们期望的结果。

图 7.5$b$ 和图 7.5$d$ 描述了模型中两个函数的斜率参数变化的影响。$b$ 的增加意味着需求曲线的斜率的绝对值增大,即它将

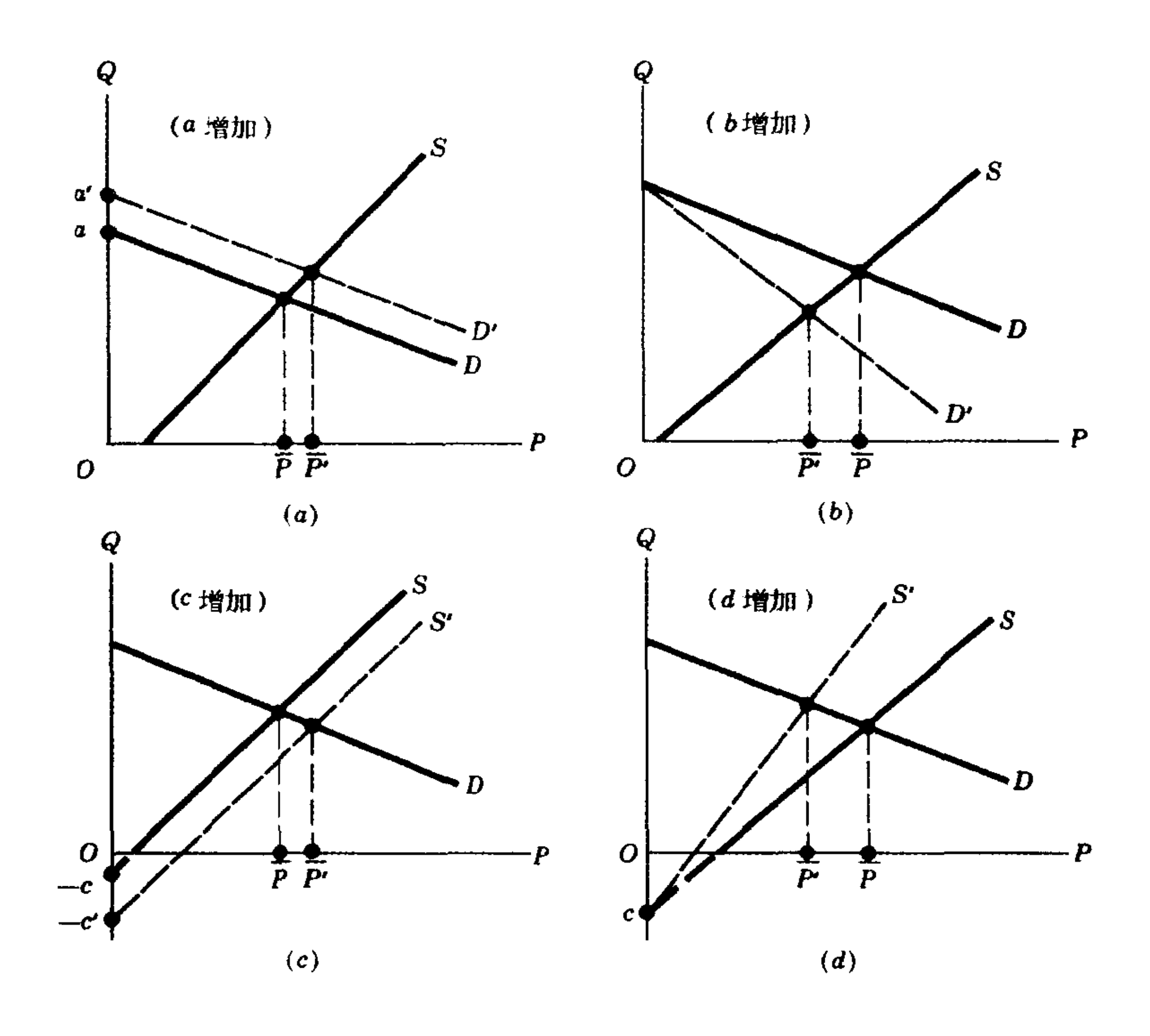



图 7.5

变得更陡。按照结论 $\partial\overline{P}/\partial b<0$,我们知道此图中的 $\overline{P}$ 下降。而 $d$ 的增加将使供给曲线更陡,从而使均衡价格上升。当然,这与比较静态导数 $\partial\overline{P}/\partial d$ 具有负号是一致的。

这样,(7.16)中的结论似乎都可由图形法得到。若如此,我们为什么要如此麻烦地学习微分法吗?原因是微分法至少有两个主要优点。首先,图形法受维数限制,而微分法则没有这种约束。甚至在内生变量和参数数量较多,使得难以通过图形法表示均衡状态时,我们仍可运用微分法来解决问题。其次,微分法可以给出更具一般性的结论。不论参数 $a$、$b$、$c$、$d$ 取何具体值,只要它们满足

181

正负号约束条件,(7.16)中的结论总是正确的。所以,此模型的比较静态分析结论实际上可应用于无数(线性)供给和需求函数的组合。比较而言,图形法仅适于需求和供给曲线的某些特殊情形;严格地讲,其分析结论源于所描述的具体函数。

上面,我们将偏导数用于简单市场模型的比较静态分析,但实际上我们仅完成了任务的一半,因为我们还需求出 $\bar{Q}$ 的比较静态导数。这个工作留给读者作为练习。

## 国民收入模型

我们把第 3 章中的简单国民收入模型扩大为具有三个内生变量 $Y$(国民收入)、$C$(消费)和 $T$(税收)的模型:

$$\begin{aligned} Y &= C + I_0 + G_0 \\ C &= \alpha + \beta(Y - T) \qquad (\alpha > 0;\ 0 < \beta < 1) \\ T &= \gamma + \delta Y \qquad (\gamma > 0;\ 0 < \delta < 1) \end{aligned} \tag{7.17}$$

方程组中的第一个方程给出了国民收入的均衡条件,而第二、三个方程则分别表示 $C$ 和 $T$ 在模型中是如何决定的。

这里必须对参数 $\alpha$、$\beta$、$\gamma$、$\delta$ 值的限制进行解释:因为即使可支配收入($Y-T$)为零,消费仍为正,所以 $\alpha$ 为正;而 $\beta$ 为一正分数,因其表示边际消费倾向;$\gamma$ 也为正值,因为即使 $Y$ 为零,政府也会有正的税收(源于税基而非收入);$\delta$ 是一个正的分数,因为它表示所得税率,因此,它不能超过 100%。外生变量 $I_0$(投资),$G_0$(政府支出)当然为非负,假设所有参数和外生变量彼此独立,所以,其中任何一个参数和外生变量被赋予新值,均不影响其它变量。

将(7.17)中的第三个方程代入第二个方程,再将所得方程代

入第一个方程，可以解模型得到 $\bar{Y}$。均衡收入（以简化式表示）为：

$$\bar{Y}=\frac{\alpha-\beta\gamma+I_0+G_0}{1-\beta+\beta\delta} \tag{7.18}$$

182 类似地，我们还可以求得内生变量 $C$ 和 $T$ 的均衡值，但我们这里集中讨论均衡收入。

由(7.18)式，可求得六个比较静态导数。其中，下述三个具有特别的政策意义：

$$\frac{\partial\bar{Y}}{\partial G_0}=\frac{1}{1-\beta+\beta\delta}>0 \tag{7.19}$$

$$\frac{\partial\bar{Y}}{\partial\gamma}=\frac{-\beta}{1-\beta+\beta\delta}<0 \tag{7.20}$$

$$\frac{\partial\bar{Y}}{\partial\delta}=\frac{-\beta(\alpha-\beta\gamma+I_0+G_0)}{(1-\beta+\beta\delta)^2}=\frac{-\beta\bar{Y}}{1-\beta+\beta\delta}<0 \qquad [\text{由}(7.18)] \tag{7.21}$$

(7.19)式给出**政府支出乘数**，因为 $\beta$ 小于 1，所以其符号为正，且 $\beta\delta$ 大于零。若给定参数 $\beta$ 和 $\delta$ 的数值，我们还可由(7.19)式求出乘数的数值。(7.20)式的导数可称作**非所得税乘数**，因为它表明 $\gamma$ 的单位变化，即政府源于非所得税的收入的单位变化，将对均衡收入产生何种影响。在本模型中乘数为负，因为(7.20)中的分母为正而分子为负。最后，(7.21)中的偏导数表示**所得税率乘数**。对任意正的均衡收入，此乘数在模型中也为负值。

## 投入—产出模型

开放投入产出模型的解是一个矩阵方程 $\bar{x}=(I-A)^{-1}d$。

若我们以 $B=[b_{ij}]$ 表示逆矩阵 $(I-A)^{-1}$，则一个三个产业的经济的解可以写成 $\bar{x}=Bd$，或者

$$\begin{bmatrix}\bar{x}_1\\ \bar{x}_2\\ \bar{x}_3\end{bmatrix}=\begin{bmatrix}b_{11} & b_{12} & b_{13}\\ b_{21} & b_{22} & b_{23}\\ b_{31} & b_{32} & b_{33}\end{bmatrix}\begin{bmatrix}d_1\\ d_2\\ d_3\end{bmatrix} \tag{7.22}$$

那么，解值 $\bar{x}_j$ 对外生的最终需求 $d_1$、$d_2$ 和 $d_3$ 的变化率为多少呢？一般答案为

$$\frac{\partial \bar{x}_j}{\partial d_k}=b_{jk} \qquad (j,k=1,2,3) \tag{7.23}$$

为看清这一点，我们乘开(7.22)式的 $Bd$，并将解表示成　183

$$\begin{bmatrix}\bar{x}_1\\ \bar{x}_2\\ \bar{x}_3\end{bmatrix}=\begin{bmatrix}b_{11}d_1+b_{12}d_2+b_{13}d_3\\ d_{21}d_1+b_{22}d_2+b_{23}d_3\\ b_{31}d_1+b_{32}d_2+b_{33}d_3\end{bmatrix}$$

在这个三个方程组成的方程组中，每个方程给出了作为外生最终需求的函数的特定解值。这些解值的偏导数将产生九个比较静态导数：

$$\begin{aligned}
&\frac{\partial \bar{x}_1}{\partial d_1}=b_{11} && \frac{\partial \bar{x}_1}{\partial d_2}=b_{12} && \frac{\partial \bar{x}_1}{\partial d_3}=b_{13}\\
&\frac{\partial \bar{x}_2}{\partial d_1}=b_{21} && \frac{\partial \bar{x}_2}{\partial d_2}=b_{22} && \frac{\partial \bar{x}_2}{\partial d_3}=b_{23}\\
&\frac{\partial \bar{x}_3}{\partial d_1}=b_{31} && \frac{\partial \bar{x}_3}{\partial d_2}=b_{32} && \frac{\partial \bar{x}_3}{\partial d_3}=b_{33}
\end{aligned} \tag{7.23'}$$

这是(7.23)式的展开形式。

分三个不同的列来读(7.23′)式,我们可以将每列中的三个导数组成一个矩阵(向量)导数:

$$\frac{\partial \bar{x}}{\partial d_1} \equiv \frac{\partial}{\partial d_1}\begin{bmatrix}\bar{x}_1\\ \bar{x}_2\\ \bar{x}_3\end{bmatrix} = \begin{bmatrix}b_{11}\\ b_{21}\\ b_{31}\end{bmatrix} \qquad \frac{\partial \bar{x}}{\partial d_2} = \begin{bmatrix}b_{12}\\ b_{22}\\ b_{32}\end{bmatrix} \qquad \frac{\partial \bar{x}}{\partial d_3} = \begin{bmatrix}b_{13}\\ b_{23}\\ b_{33}\end{bmatrix} \tag{7.23''}$$

因为(7.23″)中的三个列向量仅是矩阵 $B$ 的列,经过进一步的考虑,我们可以将九个导数概括成一个单一矩阵导数 $\partial\bar{x}/\partial d$。给定 $x = Bd$,我们能简单写成

$$\frac{\partial \bar{x}}{\partial d} = \begin{bmatrix}b_{11} & b_{12} & b_{13}\\ b_{21} & b_{22} & b_{23}\\ b_{31} & b_{32} & b_{33}\end{bmatrix} = B$$

这是这个开放的投入产出模型的全部比较静态导数的简洁表达方式。显然,此矩阵导数可轻易由现在的三个产业经济推广到一般的 $n$ 产业经济。

投入产出模型的比较静态导数是经济规划的有用工具,因为它为下述问题提供了答案:如果反映在$(d_1, d_2, \cdots, d_n)$中的规划目标被修订,且如果我们希望照顾到该经济中所有直接和间接的需要以完全避免瓶颈因素,我们需如何改变 $n$ 个产业的产出目标呢?

184 **练习 7.5**

---

1 检验(7. 15)式的均衡数量的比较静态性质,并以图形分析验证你的结论。

2 根据(7. 18)式,求出偏导数 $\partial\overline{Y}/\partial I_0$、$\partial\overline{Y}/\partial\alpha$ 以及 $\partial\overline{Y}/\partial\beta$。解

释其含义并确定其正负号。

3 数字投入产出模型（5.21）在5.7节中被解出：

（$a$）可推导出多少个比较静态导数？

（$b$）按（7.23′）和（7.23″）的形式写出这些导数。

---

## 7.6 雅可比行列式的注释

前面对偏导数的研究，完全由于比较静态分析的动机所驱使。但偏导数也提供了一种检验 $n$ 个变量 $n$ 个函数的集合中是否存在函数（线性或非线性的）相关的方法。这个问题与雅可比行列式（以雅可比的名字命名）的概念相联系。

考察两个函数

$$\begin{aligned} y_1 &= 2x_1 + 3x_2 \\ y_2 &= 4x_1^2 + 12x_1x_2 + 9x_2^2 \end{aligned} \tag{7.24}$$

若我们求出所有四个偏导数

$$\frac{\partial y_1}{\partial x_1}=2 \quad \frac{\partial y_1}{\partial x_2}=3 \quad \frac{\partial y_2}{\partial x_1}=8x_1+12x_2 \qquad \frac{\partial y_2}{\partial x_2}=12x_1+18x_2$$

并按规定顺序将其重排成一个方阵，将其称作雅可比矩阵，并以 $J$ 表示；然后取其行列式，便得到所谓的**雅可比行列式**（或简称雅可比式）以 $|J|$ 表示：

$$|J| \equiv \begin{vmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_1}{\partial x_2} \\ \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_2}{\partial x_2} \end{vmatrix} = \begin{vmatrix} 2 & 3 \\ (8x_1+12x_2) & (12x_1+18x_2) \end{vmatrix} \tag{7.25}$$

为节约空间，雅可比行列式有时也表示成

$$|J| \equiv \left|\frac{\partial\ (y_1,\ y_2)}{\partial\ (x_1,\ x_2)}\right|$$

185更一般地，若我们有具有 $n$ 个变量的 $n$ 个可微函数（不必都是线性的）：

$$\begin{aligned} y_1 &= f^1(x_1, x_2, \cdots, x_n) \\ y_2 &= f^2(x_1, x_2, \cdots, x_n) \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n &= f^n(x_1, x_2, \cdots, x_n) \end{aligned} \tag{7.26}$$

其中，$f^n$ 表示第 $n$ 个函数（而非函数的 $n$ 次幂）。我们可以由此推导出 $n^2$ 个偏导数。将其归在一起，将得到雅可比行列式

$$|J| \equiv \left|\frac{\partial(y_1, y_2, \cdots, y_n)}{\partial(x_1, x_2, \cdots, x_n)}\right| \equiv \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial y_n}{\partial x_1} & \frac{\partial y_n}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_n} \end{vmatrix} \tag{7.27}$$

下述定理给出 $n$ 个函数集合中函数相关存在性的雅可比检验：当且仅当（7. 26）式中的 $n$ 个函数 $f^1$，…，$f^n$ 为函数（线性或非线性的）相关时，（7. 27）所定义的雅可比行列式对所有的 $x_1$，…，$x_n$ 值恒为零。

作为一个例子，对于（7. 24）中的两个函数，（7. 25）所给出的雅可比行列式的值为

$$|J| = (24x_1 + 36x_2) - (24x_1 + 36x_2) = 0$$

即雅可比式对 $x_1$ 和 $x_2$ 的所有值均为零。因此，按照上述定理，

(7. 24)中的两个函数必然是函数相关的。读者可以自己验证这一点：$y_2$ 只是 $y_1$ 的平方，因此，它们确实是函数相关的，这里是非线性相关。

现在我们考察线性函数的特殊情况。我们前面已经证明，当且仅当下列线性方程组

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= d_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= d_2 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= d_n
\end{aligned}
\tag{7.28}
$$

的行列式 $|A| = 0$ 时，线性方程组系数矩阵 $A$ 的行间存在线性相关。这个结果可以看作函数相关的雅可比行列式检验准则的一个具体应用。

将 (7. 28) 式中每一方程的左边作为一个单独的包括 $n$ 个变量 $x_1$，…，$x_n$ 的函数，并将这些函数以 $y_1$，…，$y_n$ 表示。可以证明，这些函数的偏导数为 $\partial y_1/\partial x_1 = a_{11}$，$\partial y_1/\partial x_2 = a_{12}$ 等等，因而我们可以一般地写成 $\partial y_i/\partial x_j = a_{ij}$。所以，$n$ 个函数的雅可比行列式的元素恰好是已按正确顺序排列的系数矩阵 $A$ 的元素。即我们有 $|J| = |A|$，这样，$y_1$，…，$y_n$ 间的函数相关（或者，实际上是一样的，系数矩阵 $A$ 各行间的函数相关）186 的雅可比行列式检验准则在现在的线性情况下，等价于判别准则 $|A| = 0$。

前面，雅可比行列式是在含 $n$ 个变量的 $n$ 个函数系统中进行讨论的。但需要指出的是，(7. 27) 式中的雅可比行列式即使是在 (7. 26) 中的函数包含的变量多于 $n$ 个时，仍然有定义：

$$y_i = f^i(x_1, \cdots, x_n, x_{n+1}, x_{n+2}) \ (i = 1, 2, \cdots, n)$$

在此情况下，若我们将其中的任意两个变量（比如 $x_{n+1}$，$x_{n+2}$）保持不变，或者将其视为参数，我们又可以有含 $n$ 个变量的 $n$ 个函数，并形成雅可比行列式。进而，通过使不同的 $x$ 变量对保持不变，我们可以形成不同的雅可比行列式。在第 8 章讨论隐函数定理时还会遇到这种情况。

## 练习 7.6

---

1 运用雅可比行列式检验下列函数对中是否存在函数相关：

(a) $y_1 = 3x_1^2 + x_2$

$y_2 = 9x_1^4 + 6x_1^2(x_2 + 4) + x_2(x_2 + 8) + 12$

(b) $y_1 = 3x_1^2 + 2x_2^2$

$y_2 = 5x_1 + 1$

2 将三个函数 $\overline{x}_i = f^i(d_1, d_2, d_3)$ $(i = 1, 2, 3)$ 看作一个集合：

(a) 写出 3×3 雅可比行列式。它与（7.23′）是否具有某些联系？我们可以写出 $|J| = |B|$ 吗？

(b) 因为 $B \equiv (I - A)^{-1}$，我们可得出结论 $|B| \neq 0$ 吗？对（7.22）中的三个方程，我们能由此得出何结论？

---

# 第8章 一般函数模型的比较静态分析

通过上一章对偏导数的学习,我们已可以处理一些简单的比 187 较静态问题,其中模型的均衡解以简化型显式表示。在那种情况下,解的偏导数将直接给出所期望的比较静态信息。读者可以回忆一下,偏导数的定义要求自变量(比如 $x_i$)之间不存在任何函数相关,从而 $x_1$ 可独立变化而不对 $x_2$, $x_3$, $\cdots$, $x_n$ 的值产生影响。当将其应用于比较静态分析时,这意味着简化型解中的参数和(或)外生变量必定为相互无关的。因为这实际上是为建模而定义成的预定信息,其相互影响的可能性已被内在地排除了。因而上一章采用偏导数的方法是完全合理的。

但是,当模型中包含一般函数,因而不能得到简化型显式解时,便不能再指望会有如何的方便了。在此情况下,我们必须从模型原来给定的方程中直接求出比较静态导数。例如,取一个具有两个内生变量 $Y$ 和 $C$ 的简单国民收入模型:

$$Y = C + I_0 + G_0$$

$$C = C(Y, T_0) \qquad [T_0\text{:外生的税收}]$$

它可简化为一个方程(均衡条件)

$$Y = C(Y, T_0) + I_0 + G_0$$ 188

以解出 $\bar{Y}$。但由于 $C$ 函数的一般形式,没有合适的显式解。因此,我们必须由此方程直接求出比较静态导数。我们能解决这一

问题吗？我们能遇到什么特殊困难呢？

我们假设均衡解 $\bar{Y}$ 确实存在。则在非常一般的条件(后面将讨论)下，我们可取 $\bar{Y}$ 为外生变量 $I_0$、$G_0$、$T_0$ 的可微函数。因此，我们可写出方程

$$\bar{Y} = \bar{Y}(I_0, G_0, T_0)$$

尽管我们不能确定该函数所取的显式形式。进而，在均衡值 $\bar{Y}$ 的某个邻域，下列恒等式成立：

$$\bar{Y} \equiv C(\bar{Y}, T_0) + I_0 + G_0$$

这种恒等式称作均衡恒等式，因为此式只不过是均衡条件中的 $Y$ 变量被均衡值 $\bar{Y}$ 替代而已。既然 $\bar{Y}$ 已经出现，似乎初看起来，对此恒等式简单求偏导便会得到所期望的比较静态导数，比如，$\partial\bar{Y}/\partial T_0$。但不幸的是，情况并非如此。因为 $\bar{Y}$ 是 $T_0$ 的函数，$C$ 函数的两个自变量并非不相关。特别是在此情况下，$T_0$ 不仅直接影响 $C$，而且通过 $\bar{Y}$ 间接影响 $C$。因此，求偏微分不再适合于实现我们的目的。那么，我们怎样解决这一问题呢？

答案是，我们必须借助于全微分(以区别于偏微分)，基于全微分的概念，求全微分的过程可使我们得到一个相关的概念——全导数，当 $T_0$ 影响另一自变量 $\bar{Y}$ 时，它度量函数 $C(\bar{Y}, T_0)$ 对自变量 $T_0$ 的变化率。因此，一旦我们掌握了这些概念，便可以处理那些自变量并不独立的函数，这样便可以清除我们在研究一般函数模型比较静态分析中迄今所遇到的最大障碍。但在讨论这些概念之前，我们应先导入微分这一概念。

## 8.1 微分

符号 $dy/dx$，作为函数 $y = f(x)$ 的导数，迄今为止一直被视

为一个单独的整体。现在我们将其重新解释为两个量 $dy$ 和 $dx$ 的比率。

## 微分与导数

189

给定函数 $y=f(x)$，特定的 $\Delta x$ 将导致相应的变化 $\Delta y$，且我们可以使用差商 $\Delta y/\Delta x$ 表示 $y$ 对 $x$ 的变化率。因为

$$\Delta y \equiv \left(\frac{\Delta y}{\Delta x}\right)\Delta x \tag{8.1}$$

成立，一旦变化率 $\Delta y/\Delta x$ 和 $x$ 的变差已知，则可求出 $\Delta y$ 的大小。

当 $\Delta x$ 为无穷小，$\Delta y$ 也将为无穷小，差商 $\Delta y/\Delta x$ 则变成导数 $dy/dx$。然而若我们分别以 $dx$ 和 $dy$（不用 $\Delta x$ 和 $\Delta y$）表示 $x$ 和 $y$ 的无穷小变化，恒等式(8.1)变成

$$dy \equiv \left(\frac{dy}{dx}\right)dx \quad \text{或} \quad dy \equiv f'(x)dx \tag{8.2}$$

符号 $dy$ 和 $dx$ 分别称为 $y$ 和 $x$ 的微分。

将恒等式(8.2)两边被 $dx$ 除，我们有

$$\frac{(dy)}{(dx)} \equiv \left(\frac{dy}{dx}\right) \quad \text{或} \quad \frac{(dy)}{(dx)} \equiv f'(x) \tag{8.2'}$$

此结果表明导数 $(dy/dx)\equiv f'(x)$ 可以解释为两个独立微分 $dy$ 和 $dx$ 的商。

根据(8.2)，一旦我们知道函数 $y=f(x)$ 的导数，$dy$ 便可以写成 $f'(x)dx$。因此，可以将 $f'(x)$ 看作一个"转换因子"，它将无穷小变化 $dx$ 转化为相应的无穷小变化 $dy$。

**例 1** 已知 $y=3x^2+7x-5$，求 $dy$。函数的导数为 $dy/dx=6x+7$，因此所求微分为

$$dy = (6x+7)dx \tag{8.3}$$

这个结果可用于计算由给定 $x$ 的变化所引起的 $y$ 的变化。但应记住的是,微分 $dy$ 和 $dx$ 仅指无穷小变化;因此,若我们将一个很大的值 $\Delta x$ 代入(8.3),所得到的 $dy$ 仅可以作为相应的 $y$ 的确切变化 $\Delta y$ 的近似值。假设 $x$ 由 5 变至 5.01,我们由(8.3)式计算 $dy$。首先,我们令 $x=5$, $dx=0.01$,并将其代入(8.3)式,结果 $dy=37(0.01)=0.37$。这个数字与 $y$ 的实际变化相比怎么样呢?当变化前 $x=5$,我们可由已知函数计算出 $y=105$,而当 $x=5.01$时,我们得到 $y=105.3703$。因此 $y$ 的真实变化 $\Delta y=0.3703$,所以,我们的结果 $dy=0.37$ 的近似误差为 0.0003。

190

近似误差的来源可用图 8.1 来说明。对于图中给定的 $\Delta x$(距离 $AC$),$y$ 的实际变化,或 $\Delta y$,是距离 $CB$。如果我们用直线 $AB(=\Delta y/\Delta x=CB/AC)$作为相应的变化率,并应用(8.1)来求 $\Delta y$,我们将得到正确的答案

$$\Delta y=\left(\frac{\Delta y}{\Delta x}\right)\Delta x=\frac{CB}{AC}AC=CB$$

但在使用(8.3)时[它是(8.2)的特殊形式],我们实际上使用导数 $dy/dx$ 代替了 $\Delta y/\Delta x$,即我们在计算中用切线 $AD$ 的斜率$(=CD/AC)$代替了直线 $AB$ 的斜率。因此,我们得到答案:

$$dy=\left(\frac{dy}{dx}\right)\Delta x=\frac{CD}{AC}AC=CD$$

它与真实变化 $CB$ 相差误差 $DB$。当然,可以预期,随着 $\Delta x$ 趋小,点 $B$ 趋近于点 $A$,误差就会变小。

求微分 $dy$ 的过程被称作微分。回忆一下,我们前面一直将微分作为求导的同义语而使用,但却未给出适当的解释。然而,按照我们将导数解释成两个微分的商的作法,这个术语的合理性便不言自明了。但将同一术语“微分”既用于描述求微分 $dy$ 的过

程，又用于描述求导数 $dy/dx$ 的过程，总是有点不明确。为避免混淆，在取导数 $dy/dx$ 时，我们通常就把"微分"这个词说成是对 $x$ 求导。

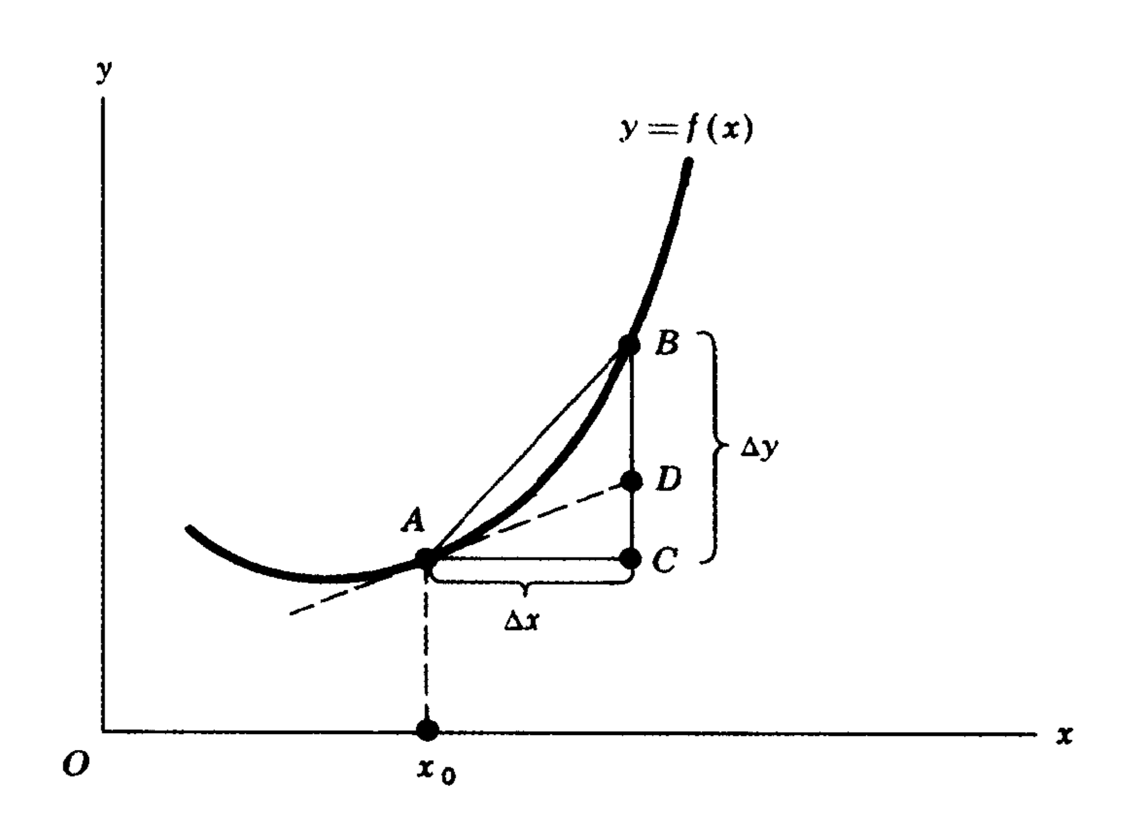



图　8.1

由(8.2)式应当很清楚，给定函数 $y=f(x)$，我们总可以(1)用 $dx$ 191 除以已知微分 $dy$，从而将其变换成导数 $dy/dx$；(2)用 $dx$ 乘以已知导数 $dy/dx$，从而将其变换成微分 $dy$。

## 微分与点弹性

为举例说明微分在经济学中的应用，我们考察函数弹性这一概念。例如，对于需求函数 $Q=f(P)$，弹性定义为 $(\Delta Q/Q)/(\Delta P/P)$。但若 $P$ 变化无穷小，表达式 $\Delta P$ 和 $\Delta Q$ 将变成 $dP$ 和 $dQ$，则此时的弹性度量便具有需求的点弹性的含义，我们以 $\varepsilon_d$(希

腊字母 *epsilon*,用于表示弹性)表示。[①]

$$\varepsilon_d \equiv \frac{dQ/Q}{dP/P} = \frac{dQ/dP}{Q/P} \tag{8.4}$$

注意上式最右边的部分,我们已将微分 $dQ$ 和 $dP$ 重排成比率 $dQ/dP$,它是需求函数 $Q=f(P)$ 的导数或边际函数,因为类似地,我们还可以将分母中的比率 $Q/P$ 看作需求函数的平均函数,所以(8.4)中的需求点弹性 $\varepsilon_d$ 可以看作是需求函数的边际函数对平均函数的比率。

实际上,刚才所描述的关系不仅对需求函数的点弹性,而且对其它函数的点弹性都是成立的,因为对任意给定总函数 $y=f(x)$,我们可以将 $y$ 对 $x$ 的点弹性公式写成

$$\varepsilon_{yx} = \frac{dy/dx}{y/x} = \frac{\text{边际函数}}{\text{平均函数}}$$

为方便计,我们用弹性的绝对值来度量函数在某一特定的点是否有弹性。例如,在需求函数的例子中,我们规定:

$$\text{当}|\varepsilon_d| \gtreqless 1\text{ 时,需求在该点}\begin{cases}\text{有弹性}\\ \text{为单位弹性}\\ \text{缺乏弹性}\end{cases}$$

**例 2** 若需求函数为 $Q=100-2P$,求 $\varepsilon_d$。给定函数的边际函数和平均函数为

$$\frac{dQ}{dP} = -2 \qquad \text{和} \qquad \frac{Q}{P} = \frac{100-2P}{P}$$

所以其比率使我们得出

---

① 点弹性还可以解释成当 $\Delta P \to 0$ 时,$\frac{\Delta Q/Q}{\Delta P/P} = \frac{\Delta Q/\Delta P}{Q/P}$ 的极限,它与(8.4)式的结果相同。

$$\Sigma_d = \frac{-P}{50-P}$$

在上式中，弹性表现为 $P$ 的函数。只要具体的价格选定，点弹性的大小便确定了。例如，当 $P=25$ 时，我们有 $\varepsilon_d=-1$，或 $|\varepsilon_d|=1$，所以该点的需求弹性为单位弹性。但当 $P=30$ 时，我们有 $|\varepsilon_d|=1.5$，因此，在此价格水平上，需求富于弹性。更一般地，可以验证，在本例中，当 $25<P<50$ 时，我们有 $|\varepsilon_d|>1$；当 $0<P<25$ 时，$|\varepsilon_d|<1$。（在这里.价格 $P>50$ 有意义吗？） 192

可能会有些离题，但仍需补充的是，将两个微分的比率看作导数，从而将函数点弹性的公式变换为边际对平均函数之比例，使得可以用图形法迅速确定点弹性。图 8.2 分别描述了曲线斜率为负和为正的两种情况。在每一情况下，在曲线上 $A$ 点或定义域中 $x=x_0$ 点的边际函数值，由切线 $AB$ 的斜率来度量。而在每一情况下的平均函数值，则由直线 $OA$（连接原点与曲线上给定点 $A$ 的直线，像一个矢径）的斜率来度量。因为在点 $A$，我们有 $y=x_0A$ 和 $x=Ox_0$，所以平均函数为 $y/x=x_0A/Ox_0=OA$ 的斜率。因而在 $A$ 点的弹性便可以通过比较上述两个斜率的数值来确定：若 $AB$ 陡于 $OA$，则函数在点 $A$ 富于弹性；若相反，则函数在 $A$ 缺乏弹性。因此，图 8.2$a$ 所绘函数在点 $A$ 或 $x=x_0$ 是缺乏弹性的，而在图8.2$b$ 中的 $A$ 点则是富于弹性的。

进一步看，所比较的两个斜率直接取决于两个角 $\theta_m$ 和 $\theta\alpha$（希腊字母 *theta*，下标 $m$ 和 $\alpha$ 分别表示边际和平均）的大小。因此，我们还可以比较两个角来代替比较两个斜率。再观察图 8.2，读者可以看到在图 $a$ 中的点 $A$，$\theta_m<\theta_\alpha$，表明边际值小于平均值，所以函数在点 $A$ 缺乏弹性。图 $b$ 中的情况恰好完全相反。

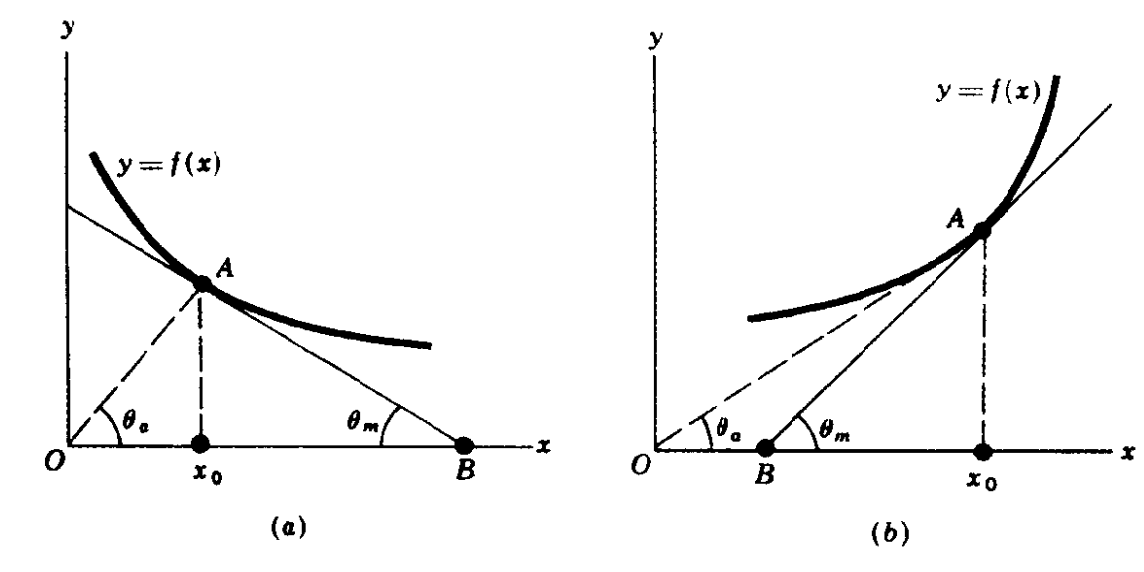



图 8.2

有时,我们对给定曲线的单位弹性点感兴趣。现在可以很容易地找到这个点。若曲线斜率为负,如图 8.3$a$ 所示,我们可以找到点 $C$,使得直线 $OC$ 和切线 $BC$ 与 $x$ 轴形成的角的大小相同,尽管方向相反。在曲线斜率为正的情况下,如图 8.3$b$ 所示,我们仅需要在曲线上找到点 $C$,使得切线切于 $C$ 点,并经适当延伸,可通过原点即可。

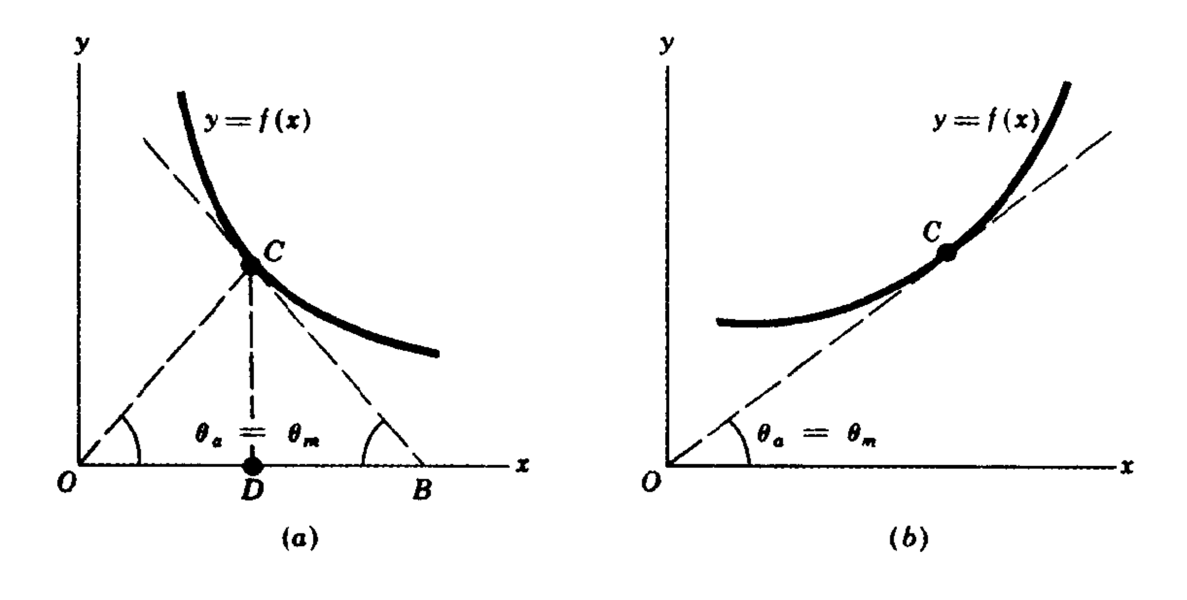



图 8.3

我们必须提醒读者，刚才所描述的图形法是以这一假设为基础的：函数 $y=f(x)$ 是把因变量 $y$ 作为纵轴而绘成的。具体来说，当我们将此方法应用于需求函数时，我们应使 $Q$ 为纵轴。(假设 $Q$ 实际上在作图时被绘成横轴，我们读点弹性的方法应作何种修正呢?)

### 练习 8.1

---

1 已知下列函数，求微分 $dy$:

(a) $y=-x(x^2+3)$ (b) $y=(x-8)(7x+5)$ (c) $y=\frac{x}{x^2+1}$

2 给定进口函数 $M=f(Y)$，其中 $M$ 为进口，$Y$ 为国民收入。按进口倾向表示进口的收入弹性 $\varepsilon_{MY}$。

3 给定消费函数 $C=a+bY(a>0;0<b<1)$:

(a) 求其边际函数和平均函数。

(b) 求消费的收入弹性 $\varepsilon_{CY}$，假设 $Y>0$，确定 $\varepsilon_{CY}$ 的符号。

(c) 证明此消费函数在所有正的收入水平上均是缺乏弹性的。

4 给定 $Q=k/P^n$，其中 $k$ 与 $n$ 为正常数，

(a) 在这种情况下，弹性取决于价格吗?

(b) 在 $n=1$ 的特定情况下，需求曲线是何形状? 需求的点弹性为多少?

---

194

## 8.2 全微分

微分的概念很容易推广至具有两个或更多自变量的函数的情况。考察储蓄函数

$$S = S(Y, i) \tag{8.6}$$

其中 $S$ 为储蓄，$Y$ 为国民收入，$i$ 为利息率。像这里要使用的所有函数均将采用的假设一样，我们假设此函数连续，且具有连续的(偏)导数，这实际上是函数平滑且处处可导的另一种表述方式。我们知道，偏导数 $\partial S/\partial Y$(或 $S_Y$)度量 $S$ 对 $Y$ 的无穷小变化的变化率，或者简言之，它表示边际储蓄倾向。因此，由于 $Y$ 的变化所引致的 $S$ 的变化可以表达成$(\partial S/\partial Y)dY$ 表示，此式与(8.2)式右边的表达式具有可比性。基于同样的理由，由 $i$ 的无穷小变化引致的 $S$ 变化可表示成$(\partial S/\partial i)di$。则 $S$ 的总变化等于

$$dS = \frac{\partial S}{\partial Y}dY + \frac{\partial S}{\partial i}di$$

或用另一种记号：

$$dS = S_Y dY + S_i di$$

注意，两个偏导数 $S_Y$ 和 $Si$ 同样起“变换因子”的作用：它将无穷小变化 $dY$ 和 $di$ 分别变换为对应的变化 $dS$。而表达式 $dS$，作为两种不同原因所引致的变化之和，称作储蓄函数的**全微分**。求此全微分的过程，称作**全微分法**。

当然，尽管 $i$ 保持不变时，$Y$ 仍可变化。在此情况下，$di = 0$，则全微分将简化为**偏微分** $dS = (\partial S/\partial Y)dY$。两边同时除以 $dY$，我们得到：

$$\frac{\partial S}{\partial Y} = \left(\frac{dS}{dY}\right)_{i\text{不变}}$$

因此，很明显，偏导数 $\partial S/\partial Y$ 也可以解释成两个微分 $dS$ 与 $dY$ 的比率，其前提是函数中另一自变量 $i$ 保持不变。依照完全类似的

195 方式，当 $dY = 0$ 时，我们还可形成另一个偏微分 $dS = (\partial S/\partial i)$，

并可将偏导数 $\partial S/\partial i$ 解释为微分 $dS$($Y$ 保持不变)对微分 $di$ 的比率,注意,$dS$ 和 $di$ 现在均可代表独立的微分,但表达式 $\partial S/\partial i$ 仍然是一个整体。

具有 $n$ 个自变量的更一般情况的例子可由一般形式的效用函数给出:

$$U = U(x_1, x_2, \cdots, x_n) \tag{8.7}$$

此函数的全微分可以表示成

$$dU = \frac{\partial U}{\partial x_1}dx_1 + \frac{\partial U}{\partial x_2}dx_2 + \cdots + \frac{\partial U}{\partial x_n}dx_n$$

或

$$dU = U_1\,dx_1 + U_2\,dx_2 + \cdots + U_n dx_n = \sum_{i=1}^{n} U_i dx_i$$

其中,右边表达式中的每一项表示由某一自变量无穷小变化引致的$U$ 的变化量。上式中第一项 $U_1 dx_1$ 的经济意义是第一个商品的边际效用乘以该商品消费的增量。其它各项也类似。因此,这些项的和表示由所有可能的原因所引致的效用变化的总和。

像任意其它函数一样,储蓄函数(8.6)和效用函数(8.7)均可产生类似于(8.5)所定义的弹性量度。但在这两种情况下,弹性量度必须仅根据一个自变量的变化来定义,因此,这个储蓄函数有两个这样的弹性量度,效用函数则有 $n$ 个这样的弹性量度。它们被称作偏弹性。对储蓄函数而言,偏弹性可以表示成

$$\varepsilon_{SY} = \frac{\partial S/\partial Y}{S/Y} = \frac{\partial S}{\partial Y}\frac{Y}{S} \text{ 和 } \varepsilon_{Si} = \frac{\partial S/\partial i}{S/i} = \frac{\partial S}{\partial i}\frac{i}{S}$$

对于效用函数,$n$ 个偏弹性可以简洁地表示如下:

$$\varepsilon_{Ux_i} = \frac{\partial U}{\partial x_i}\frac{x_i}{U} \quad (i = 1, 2, \cdots, n)$$

## 练习 8.2

---

1 求下列函数的全微分：

(a) $z=3x^2+xy-2y^3$ (b) $U=2x_1+9x_1x_2+x_2^2$

2 求下列函数的全微分：

(a) $y=\dfrac{x_1}{x_1+x_2}$ (b) $y=\dfrac{2x_1x_2}{x_1+x_2}$

3 某商品的供给函数为

$Q=a+bP^2+R^{1/2}$ $(a<0,\ b>0)$[$R$:降雨量]：

求供给的价格弹性 $\varepsilon_{QP}$，供给的降雨量弹性 $\varepsilon_{QR}$。

196 4 上题中两个偏弹性如何随 $P$ 和 $R$ 的变化而变化？若 $P$ 和 $R$ 为正，变化是单调的吗？

5 外国对我国出口的需求 $X$ 取决于外国的收入 $Y_f$，我国的价格水平 $P$：$X=Y_f^{1/2}+P^{-2}$。求：外国对我国出口的需求相对于我国价格水平的偏弹性。

---

# 8.3 微分法则

给定函数

$$y=f(x_1,\ x_2)$$

求全微分 $dy$ 的一个直接方法是求偏导数 $f_1$ 和 $f_2$，并将其代入方程

$$dy=f_1dx_1+f_2\,dx_2$$

但有时应用某些微分法则可能会更方便；由于微分法则与前面学习的导数公式有着极大的相似之处，所以记忆起来非常方便。

令 $k$ 为常数，$u$ 和 $v$ 为变量 $x_1$ 和 $x_2$ 的两个函数。则下列法

则成立：[1]

**法则 I** $dk=0$ （参阅常函数导数）

**法则 II** $d(cu^n)=cnu^{n-1}du$ （参见幂函数微分法则）

**法则 III** $d(u\pm v)=du\pm dv$ （参见和差微分法则）

**法则 IV** $d(uv)=v\,du+u\,dv$ （参见积的微分法则）

**法则 V** $d\left(\dfrac{u}{v}\right)=\dfrac{1}{v^2}(v\,du-u\,dv)$ （参见商的微分法则）

这里我们不证明这些法则，而仅说明其实际应用。

**例 1** 求函数

$$y=5x_1^2+3x_2$$

的全微分。直接法要求计算偏导数 $f_1=10x_1$，$f_2=3$，然后可以写出

$$dy=f_1dx_1+f_2dx_2=10x_1dx_1+3dx_2$$

但我们还可以令 $u=5x_1^2$，$v=3x_2$，应用上述法则可得到如下同样的答案： 197

$$\begin{aligned} dy &= d(5x_1^2)+d(3x_2) \qquad [\text{由法则 III}] \\ &= 10x_1dx_1+3dx_2 \qquad [\text{由法则 II}] \end{aligned}$$

**例 2** 求下列函数的全微分：

$$y=3x_1^2+x_1x_2^2$$

因 $f_1=6x_1+x_2^2$，$f_2=2x_1x_2$，则所求微分为

$$dy=(6x_1+x_2^2)dx_1+2x_1x_2dx_2$$

通过应用已知法则，也可得出如下同样的结果

---

① 当 $u$ 和 $v$ 本身也是自变量，而不是其它变量 $x_1$ 和 $x_2$ 的函数时，本节所讨论的这些微分法则同样也可以应用。

$$dy = d(3x_1^2) + d(x_1 x_2^2) \qquad [由法则 III]$$

$$= 6x_1 dx_1 + x_2^2 dx_1 + x_1 d(x_2^2) \qquad [由法则 II 和 IV]$$

$$= (6x_1 + x_2^2) dx_1 + 2x_1 x_2 dx_2 \qquad [由法则 II]$$

**例 3** 求函数

$$y = \frac{x_1 + x_2}{2x_1^2}$$

的全微分。由于在本例中,偏导数为

$$f_1 = \frac{-(x_1 + 2x_2)}{2x_1^3} \quad 和 \quad f_2 = \frac{1}{2x_1^2}$$

(请读者对其进行验证以作为练习),所求全微分为

$$dy = \frac{-(x_1 + 2x_2)}{2x_1^3} dx_1 + \frac{1}{2x_1^2} dx_2$$

但通过应用微分法则,也可得到同样结果:

$$dy = \frac{1}{4x_1^4}[2x_1^2 d(x_1 + x_2) - (x_1 + x_2) d(2x_1^2)]$$

[由法则 V]

$$= \frac{1}{4x_1^4}[2x_1^2(dx_1 + dx_2) - (x_1 + x_2)4x_1 dx_1]$$

[由法则 III 和 II]

$$= \frac{1}{4x_1^4}[-2x_1(x_1 + 2x_2) dx_1 + 2x_1^2 dx_2]$$

$$= \frac{-(x_1 + 2x_2)}{2x_1^3} dx_1 + \frac{1}{2x_1^2} dx_2$$

当然,这些微分法则还可以推广到 $x_1$、$x_2$ 的两个以上函数的

198 情况。具体而言,我们可以增加下述两个法则:

**法则 VI** $\quad d(u \pm v \pm w) = du \pm dv \pm dw$

**法则 VII**　　$d(uvw) = vw\,du + uw\,dv + uv\,dw$

为导出法则 VII,我们可以应用熟悉的技巧,令 $z = vw$,从而

$d(uvw) = d(uz) = zdu + udz$　[由法则 IV]

则,再对 $dz$ 应用法则 IV,我们得到中间结果

$$dz = d(vw) = wdv + vdw$$

将其代入上一个方程,则得到所求的最终结果:

$d(uvw) = vw\,du + u(w\,dv + v\,dw) = vw\,du + uw\,dv + uv\,dw$

类似的步骤可用于推导法则 VI。

## 练习 8.3

---

1　运用微分法则求下列函数的全微分$(a)\,z = 3x^2 + xy - 2y^3$ 和$(b)\,U = 2x_1 + 9x_1x_2 + x_2^2$ 根据练习 8.2－1 的答案对你的答案进行验证。

2　用微分法则求下列函数的 $dy$:

$(a)\ y = \dfrac{x_1}{x_1 + x_2}$　　$(b)\ y = \dfrac{2x_1x_2}{x_1 + x_2}$

3　给定 $y = 3x_1(2x_2 - 1)(x_3 + 5)$

$(a)$ 用法则 VII 求 $dy$。

$(b)$ 若 $dx_2 = dx_3 = 0$,求 $y$ 的偏导数。

$(c)$ 由上述结果推导偏导数 $\partial y/\partial x_1$。

4　假设 $u$ 和 $v$ 为自变量而非其它变量的函数,证明法则 II、III、IV 和 V。

---

## 8.4　全导数

掌握了微分的概念,我们现在便可以回答本章开头所提出的

问题，即当 $\overline{Y}$ 与 $T_0$ 相关时，如何求函数 $C(\overline{Y}, T_0)$ 对 $T_0$ 的变化率。如前所述，答案在于全导数的概念。与偏导数概念不同，全导数的概念并不要求当 $T_0$ 变化时，自变量 $\overline{Y}$ 保持不变，这样，便可以考虑两个自变量间所假设的关系。

199

## 求全导数

为在一个更具一般性的框架内展开讨论，我们考察具有三个彼此相关（如图 8.4 所示）的变量 $y$、$x$、$w$ 的任意函数

$$y = f(x, w) \quad 其中\ x = g(w) \tag{8.8}$$

我们将图 8.4 称为通道图。由此图可清楚地看到，本例中变化的基本根源 $w$，可以通过两条渠道影响 $y$：(1)通过函数 $g$ 进而通过 $f$（加直线箭头）间接影响；(2) 通过函数 $f$（曲线箭头）直接影响。偏导数 $f_w$ 仅能反映出直接影响，而全导数才可以反映出这两种影响的联合效果。

为求得全导数，我们先对 $y$ 全微分，以得到全微分 $dy = f_x dx + f_w dw$。此方程两边同时被 $dw$ 相除，结果是

$$\begin{aligned} \frac{dy}{dw} &= f_x \frac{dx}{dw} + f_w \frac{dw}{dw} \\ &= \frac{\partial y}{\partial x} \frac{dx}{dw} + \frac{\partial y}{\partial w} \quad \left[\frac{dw}{dw} = 1\right] \end{aligned} \tag{8.9}$$

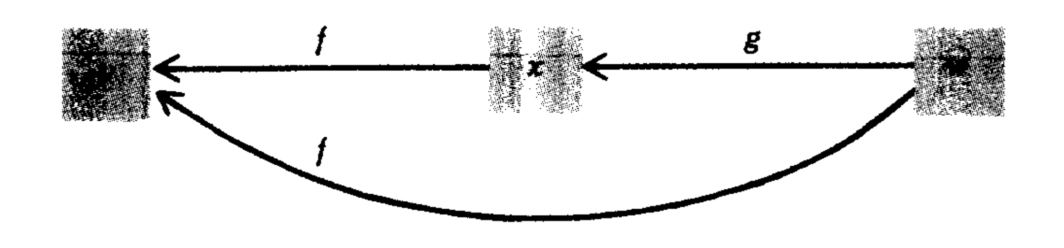



图 8.4

因为两个微分的比率可以看作一个导数，上式左边的表达式 $dy/dw$ 可以视为 $y$ 对 $w$ 变化率的度量。进一步看，若(8.9)式右边的两项能够分别等于 $w$ 对 $y$ 的间接和直接影响，则 $dy/dw$ 将确实是我们所寻求的全导数。现在，我们已经知道第二项($\partial y/\partial w$)度量直接影响，且它对应于图 8.4 中的曲线箭头。在如下箭头的帮助下[①]

$$\frac{\partial y}{\partial x} \quad \frac{dx}{dw}$$

我们对第一项 $\left(\frac{\partial y}{\partial x} \quad \frac{dx}{dw}\right)$ 进行分析可以知道，它显然度量间接影响。$w$ 的变化，即 $dw$ 首先传递到变量 $x$，并通过引致的 $x$ 的变化即 $dx$，再传递到变量 $y$。所以这确实是图 8.4 直箭头所描述的间接影响。因此，表达式(8.9)确实表示所求的全导数。求全导数 $dy/dw$ 的方法被称作 $y$ 对 $w$ 的**全微分法**。 200

**例 1** 已知函数

$$y=f(x,w)=3x-w^2 \quad \text{其中} \quad x=g(w)=2w^2+w+4$$

求全导数 $dy/dw$。由(8.9)式，全导数应为

$$\frac{dy}{dw}=3(4w+1)+(-2w)=10w+3$$

作为检验，我们可将函数 $g$ 代入函数 $f$，得到

$$y=3(2w^2+w+4)-w^2=5w^2+3w+12$$

此式仅为 $w$ 的函数。其导数 $dy/dw$ 很容易地求得为 $10w+3$，与上述结果一致。

---

① 表达式 $\frac{\partial y}{\partial x}\frac{dx}{dw}$ 易使我们联想到前面讨论过的链式法则(复合函数求导法则)，只是因为 $f$ 恰好是多元函数，所以这里出现了偏导数。

**例 2**　若我们有效用函数 $U=U(c,s)$，其中 $c$ 为咖啡的消费量，$s$ 为糖的消费量，另一个函数 $s=g(c)$ 表示两物品之间存在互补关系，则我们可以简单写成

$$U=[c,g(c)]$$

对此，我们有

$$\frac{dU}{dc}=\frac{\partial U}{\partial c}+\frac{\partial U}{\partial g(c)}g'(c)$$

## 论题的一个引申

当我们有

$$y=f(x_1,x_2,w) \qquad \text{其中}\begin{cases}x_1=g(w)\\ x_2=h(w)\end{cases} \tag{8.10}$$

情况便稍有些复杂。这种情况的通道图如图 8.5 所示：变量 $w$ 可通过三条渠道影响 $y$：(1) 通过函数 $g$，进而通过 $f$ 间接影响；(2) 通过函数 $h$，进而通过 $f$ 间接影响；(3) 通过函数 $f$ 直接影响。根

201

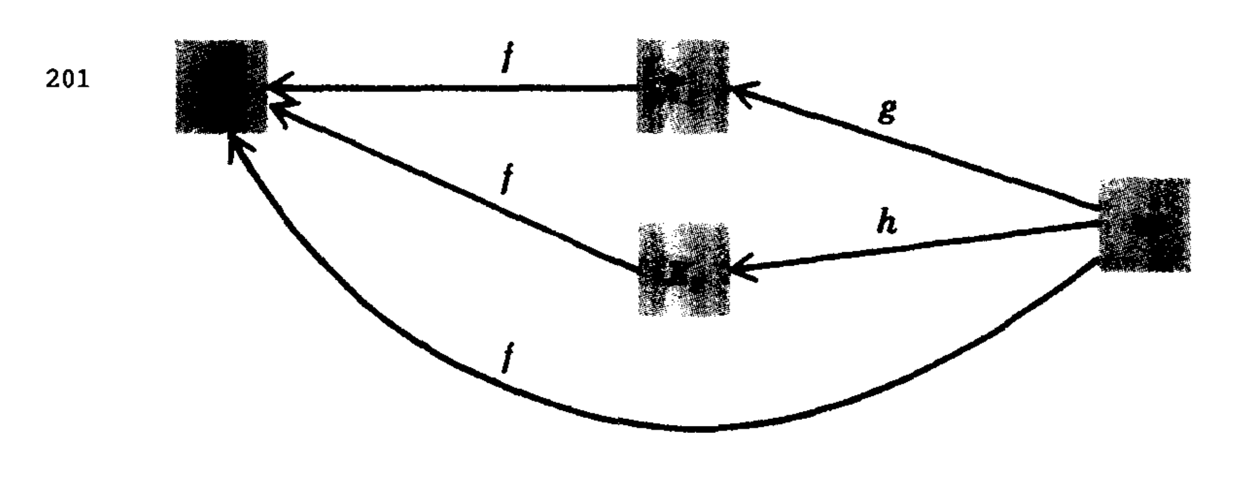



图　8.5

据前面的经验，可以预期，这三种影响分别可以表示成为 $\frac{\partial y}{\partial x_1}\frac{dx_1}{dw}$，$\frac{\partial y}{\partial x_2}\cdot\frac{dx_2}{dw}$ 和 $\frac{\partial y}{\partial w}$。这一预期是正确的，因为当我们取 $y$ 的

全微分，并将其两边分别除以 $dw$，我们得到：

$$\begin{aligned}\frac{dy}{dw} &= f_1\frac{dx_1}{dw}+f_2\frac{dx_2}{dw}+f_w\frac{dw}{dw}\\ &= \frac{\partial y}{\partial x_1}\frac{dx_1}{dw}+\frac{\partial y}{\partial x_2}\frac{dx_2}{dw}+\frac{\partial y}{\partial w}\end{aligned}\tag{8.11}$$

此式与(8.9)式是类似的。

**例 3** 令生产函数为

$$Q=Q(K,L,t)$$

其中除了两种投入 $K$ 和 $L$ 之外，还有第三个自变数 $t$，表示时间。$t$ 变量的存在表明，生产函数可能随时间而移动以反映技术变化。因此，这是一个动态的，而非静态的生产函数。因为资本与劳动也随时间而变化，所以，我们可以写成

$$K=K(t) \quad 和 \quad L=L(t)$$

则按照(8.11)的全导数公式，产出对时间的变化率可以表示成

$$\frac{dQ}{dt}=\frac{\partial Q}{\partial K}\frac{dK}{dt}+\frac{\partial Q}{\partial L}\frac{dL}{dt}+\frac{\partial Q}{\partial t}$$

或者用另一种符号表示：

$$\frac{dQ}{dt}=Q_K K'(t)+Q_L L'(t)+Q_t$$

## 论题的另一个引申

当(8.10)式变化的基本根源 $w$ 为两个共同存在的根源 $u$ 和 $v$ 所替代时，便产生以下情形：

$$y=f(x_1,x_2,u,v) \qquad 其中 \qquad \begin{cases}x_1=g(u,v)\\ x_2=h(u,v)\end{cases}\tag{8.12}$$

尽管现在的通道图可能包含更多的箭号，但其构造的基本原理是 202 相同的，因此，我们留给读者自己去完成。为求出 $y$ 对 $u$（$v$ 保持不变）的全导数，我们可以再一次取 $y$ 的全微分，并将各项通除以 $du$，得到结果

$$\frac{dy}{du}=\frac{\partial y}{\partial x_1}\frac{dx_1}{du}+\frac{\partial y}{\partial x_2}\frac{dx_2}{du}+\frac{\partial y}{\partial u}\frac{du}{du}+\frac{\partial y}{\partial v}\frac{dv}{du}$$

$$=\frac{\partial y}{\partial x_1}\frac{dx_1}{du}+\frac{\partial y}{\partial x_2}\frac{dx_2}{du}+\frac{\partial y}{\partial u}\qquad\left[\frac{dv}{du}=0\text{ 因 }v\text{ 保持不变}\right]$$

由于我们在变动 $u$ 时，使 $v$ 保持不变（因为单一导数不能处理 $u$ 和 $v$ 同时变化的情况），所以，上述结论必须以下述两种方式进行修正：(1) 右边的导数 $dx_1/du$ 和 $dx_2/du$ 应改写成偏导数符号 $\partial x_1/\partial u$ 和 $\partial x_2/\partial u$，这与(8.12)式的函数 $g$ 和 $h$ 是一致的；(2) 左边的比率 $dy/du$ 也应解释为偏导数，尽管它是由 $y$ 的全微分推导而来的，实质上仍是一个全导数。因此，为使其与简单偏导数 $\partial y/\partial u$ 相区别，我们称其为**偏全导数**，并以符号 $\S y/\S u$（以 $\S$ 而不是 $\partial$）表示；如前面的结果所示，简单偏导数只是加和起来，构成偏全导数三项中的一项。[①]

根据上述修正，我们的结果变成

$$\frac{\S y}{\S u}=\frac{\partial y}{\partial x_1}\frac{\partial x_1}{\partial u}+\frac{\partial y}{\partial x_2}\frac{\partial x_2}{\partial u}+\frac{\partial y}{\partial u}\tag{8.13}$$

它与(8.11)类似。注意，在上式右边已出现符号 $\partial y/\partial u$，从而有必要在左边采用新的符号 $\S y/\S u$ 以表示偏全导数这一含义更

① 表示偏全导数的另一种方法是

$$\left.\frac{dy}{du}\right|_{v\text{ 不变}}\text{ 或 }\left.\frac{dy}{du}\right|_{dv=0}$$

广的概念。我们还可以用完全类似的方式,推导出另一个偏全导数 $\S y/\S v$。但由于在(8.12)中 $u$ 和 $v$ 的作用完全对称,所以,我们可采用另一种更简单的方法。即将(8.13)中所有的符号 $u$ 用 $v$ 替换,便可得到 $\S y/\S v$。

运用新符号 $\S y/\S u$ 和 $\S y/\S v$ 表示偏全导数尽管有些不便,但确实可避免与(8.12)式函数 $f$ 所产生的简单偏导数 $\partial y/\partial u$ 和 $\partial y/\partial v$ 之间的混淆。然而,在 $f$ 函数取 $y=f(x_1,x_2)$ 而不用自变量 $u$ 和 $v$ 的具体情况下,简单偏导数 $\partial y/\partial u$ 和 $\partial y/\partial v$ 没有定义。因此,在这种情况下使用后一种符号表示 $y$ 对 $u$ 和 $v$ 的微全 203 导数,并没有什么不妥之处,也不可能引起什么混淆。但即使在这种情况下,使用特殊符号仍是明智之举,因为这可以增加透明度。

### 本节综述

关于全导数和全微分,我们得出三个结论:

1　在本节所讨论的各种例子和情形中,无不涉及到一个变量与第二个变量函数相关,而第二个变量又与第三个变量函数相关的情况。因而,必然出现链的概念。全导数的项中出现两个导数的一个积(或多个积)便足以证明这一点。所以,(8.9)、(8.11)、(8.13)中的全导数公式也可以视为链微分法则,或复合函数的微分法则的表达式,但它是 7.3 节所引入的链微分法则的更复杂形式。

2　导数的链并不仅限于两个“环”(两个导数相乘);全导数的概念可拓展至复合函数具有两个或多个环的情况。

3　在讨论的所有情形中,全导数,包括偏全导数,度量因变量对链中某些基本变量,或者说对某些具有外生性的或不能表示成其它变量的函数的变量的变化率。全导数和全微分法的实质,是

考虑到直接和间接的所有渠道;通过这些渠道,基本变量变化的影响能够传递到所研究的特定因变量中。

## 练习 8.4

---

1 已知下列函数,求其全导数:

(a) $z=f(x,y)=2x+xy-y^2$,其中 $x=g(y)=3y^2$

(b) $z=6x^2-3xy+2y^2$, 其中 $x=1/y$

(c) $z=(x+y)(x-2y)$, 其中 $x=2-7y$

2 给定下列函数,求其全导数:

(a) $z=x^2-8xy-y^3$, 其中 $x=3t$ 和 $y=1-t$

(b) $z=3u+vt$, 其中 $u=2t^2$和 $v=t+1$

(c) $z=f(x,y,t)$, 其中 $x=a+bt$ 和 $y=c+dt$

204

3 若生产函数为 $Q=A(t)\ K^{\alpha}L^{\beta}$,其中 $A(t)$是 $t$ 的增函数,且 $K=K_0+at$,$L=L_0+bt$,求产出对时间的变化率。

4 给定下列函数,求偏全导数 § $W$/§ $u$ 和 § $W$/§ $v$:

(a) $W=ax^2+bxy+cu$, 其中 $x=\alpha u+\beta v$ 和 $y=\gamma u$

(b) $W=f(x_1,x_2)$, 其中 $x_1=5u^2+3v$ 和 $x_2=u-4v^3$

5 画出适于(8.12)式的通道图。

6 由(8.12)式,通过取 $y$ 的全微分,然后以 $dv$ 通除 $y$ 的全微分,以正式推导出 § $y$/§ $v$ 的表达式。

---

# 8.5 隐函数的导数

全导数的概念还可以使我们求出所谓的“隐函数”的导数。

## 隐函数

一个形式为 $y=f(x)$的函数，比如

$$y = f(x) = 3x^4 \tag{8.14}$$

被称作显函数，因为变量 $y$ 被明确地表示成 $x$ 的函数。若此函数被写成另一种等价形式

$$y - 3x^4 = 0 \tag{8.14'}$$

则我们便不再有显函数了。确切地说，函数(8.14)被隐含地定义于方程(8.14′)中了。因此，当我们有形式为(8.14′)的方程，它隐含着函数 $y=f(x)$，但我们甚至不知道此函数的具体形式，我们便称其为隐函数。

因为方程的左边是两个变量 $y$ 和 $x$ 的函数，所以一般地，形式为(8.14′)的方程可以写成这种形式：$F(y,x)=0$。注意，这里我们使用大写字母 $F$，以区别于函数 $f$。函数 $F$，代表(8.14′)左边的表达式，具有两个自变数 $y$ 和 $x$。函数 $f$，代表隐函数，仅有一个自变量 $x$。当然，$F$ 函数也可以有不止两个自变量。例如，我们可以有方程 $F(y,x_1,\cdots,x_m)=0$。这样的方程可能定义成隐函数 $y=f(x_1,\cdots,x_m)$。

在上一段最后一个句子中，我们使用了一个不是很确定的词"可能"，这是经过谨慎斟酌的。因为虽然一个显函数，比如 $y=f(x)$，简单地将 $f(x)$式移至等号左边，使总可以将其变换成方程 205 $F(y,x)=0$；但这种变换并不总是可逆的。事实上，在某些情况下，给定方程 $F(y,x)=0$ 并不能定义成显函数 $y=f(x)$。例如，方程 $x^2+y^2=0$ 仅在原点(0,0)成立，因此，并不能得到我们所说的有意义的函数。另一个例子是，方程

$$F(y,x) = x^2 + y^2 - 9 = 0 \tag{8.15}$$

并不是一个函数,而仅代表一种关系,因为如图 8.6 所示,(8.15)绘出图时是一个圆,所以每一 $x$ 值并不对应一个唯一的 $y$ 值。但注意,若我们将 $y$ 值限定为非负,则我们仅得到上半个圆,它确实构成一个函数,即 $y = +\sqrt{9-x^2}$。类似地,当 $y$ 值为非负时,下半圆构成了另一个函数,$y = -\sqrt{9-x^2}$。相反,左半圆和右半圆则不能构成任何函数。

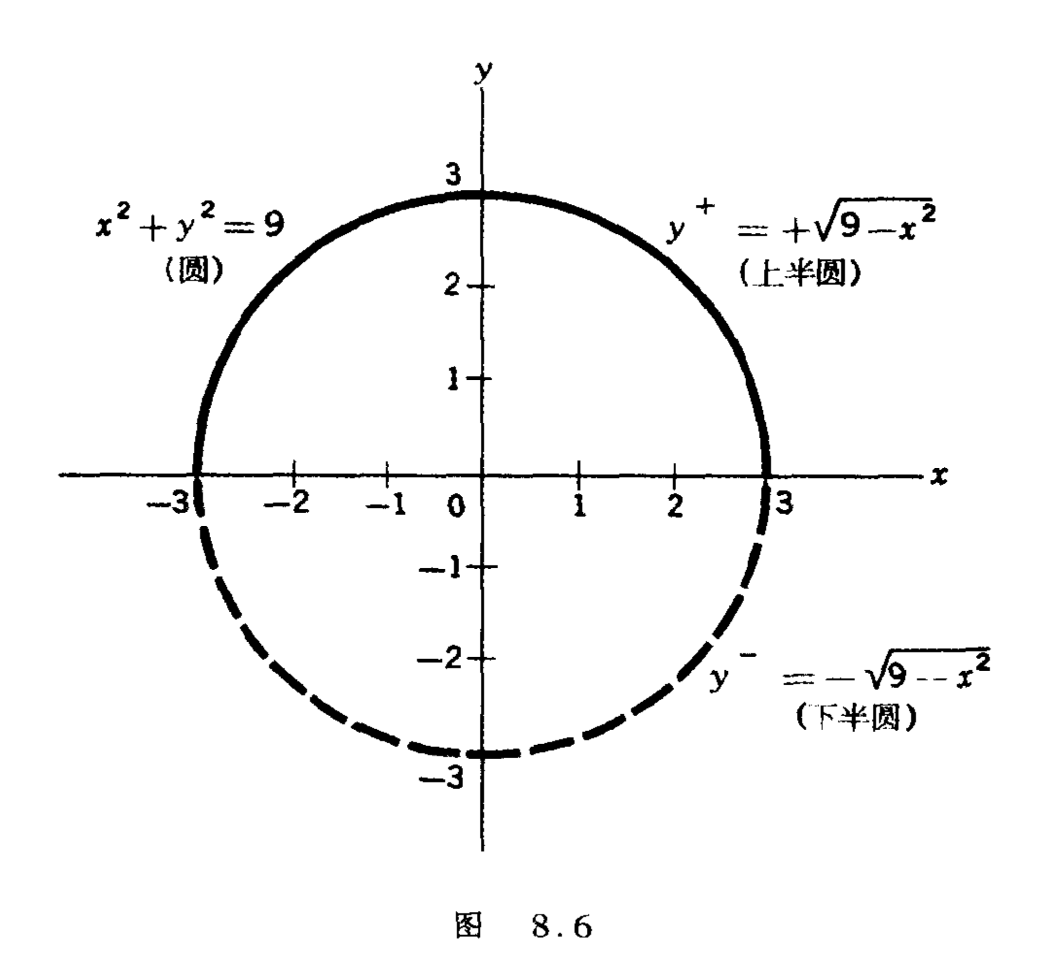



图 8.6

鉴于这种不确定性,有必要提出下述问题:是否存在已知的一般条件,在此条件下我们可以确定,给定下述形式的一般方程

$$F(y, x_1, \cdots, x_m) = 0 \tag{8.16}$$

确实可以定义一个隐函数

$$y = f(x_1, \cdots, x_m) \qquad (8.17)$$

此问题的答案由所谓的“隐函数定理”给出。该定理可表述如下：

> 给定方程(8.16)，若(a)函数 $F$ 具有连续偏导数 $F_y, F_1, \cdots, F_m$，且若(b)在点$(y_0, x_{10}, \cdots, x_{m0})$满足方程(8.16)，$F_y \neq 0$，则存在一个$(x_{10}, \cdots, x_{m0})$的 $m$ 维邻域 $N$，在此邻域中，$y$ 是变量 $x_1, \cdots, x_m$ 隐性定义的函数，其函数形式为(8.17)。此隐函数满足 $y_0 = f(x_{10}, \cdots, x_{m0})$，且对邻域 $N$ 内的每个有序 $m$ 元素组$(x_1, \cdots, x_m)$，也满足方程(8.16)——因而在此领域内使得(8.16)成为恒等式。而且，隐函数 $f$ 连续，且有连续偏导数 $f_1, \cdots, f_m$。

206

我们将此定理应用于仅含一个变量的圆的方程(8.15)。首先，我们可证明 $F_y = 2y$ 和 $F_x = 2x$ 连续。其次，我们注意到除了 $y = 0$，即除了在圆上最左边的点$(-3, 0)$和最右边的点$(3, 0)$以外，$F_y \neq 0$。因此，在圆上除$(-3, 0)$和$(3, 0)$外的任意一点，我们都可构造一个邻域，在此邻域内，方程(8.15)定义一个隐函数 $y = f(x)$。这一结论在图 8.6 中很易得到证明。在图 8.6 中除$(-3, 0)$和$(3, 0)$外的任意一点附近，我们确实可以画出一个长方形，使得封闭起来的圆周部分构成一个函数的图形，在此长方形中每个 $x$ 值，对应一个唯一的 $y$ 值。

关于隐函数定理，有以下几点值得注意。首先，定理中所列举的条件实质上是充分条件而非必要条件。这意味着如果我们在满足方程(8.16)的某点求得 $F_y = 0$，我们不能依此定理排除在该点附近隐函数的存在。因为这样的隐函数实际上是可能存在的(参

见练习 8.5—4)。[1] 其次,即便隐函数 $f$ 确实存在,定理并未给出隐函数 $f$ 的形式的任何线索。而且,该定理也并未告诉我们定义隐函数的邻域 $N$ 的大小。但尽管有这些局限性,此定理依然具有重要意义。因为当定理的条件得到满足时,讨论和利用诸如(8.17)这样的函数就变得有意义,即使我们的模型中包含方程(8.16),很难或者不可能将其求解,以便显化成 $y$ 对变量 $x$ 的函数。另外,由于定理也保证偏导数 $f_1,\cdots,f_m$ 的存在,所以现在讨论隐函数的导数也有意义了。

## 隐函数的导数

若方程 $F(y,x_1,\cdots,x_m)=0$ 对 $y$ 可解,则我们可以将其写成显函数 $y=f(x_1,\cdots,x_m)$,运用前面学过的方法求其导数。例如,解(8.15)可以得到两个独立的方程:

207

$$y^+=+\sqrt{9-x^2} \qquad [\text{上半圆}]$$

$$y^-=-\sqrt{9-x^2} \qquad [\text{下半圆}] \qquad (8.15')$$

其导数如下:

$$\frac{dy^+}{dx}=\frac{d}{dx}(9-x^2)^{\frac{1}{2}}=\frac{1}{2}(9-x^2)^{-\frac{1}{2}}(-2x)$$

$$=\frac{-x}{\sqrt{9-x^2}}=\frac{-x}{y^+} \quad (y^+\neq 0)$$

$$\frac{dy^-}{dx}=\frac{d}{dx}[-(9-x^2)^{\frac{1}{2}}]=-\frac{1}{2}(9-x^2)^{-\frac{1}{2}}(-2x)$$

---

① 另一方面,在整个邻域内若 $F_y=0$,则可得出在该邻域内不存在隐函数的结论。同理,若 $Fy=0$ 恒成立,则在任何一点均不存在隐函数。

$$= \frac{x}{\sqrt{9 - x^2}} = \frac{-x}{y^-} \quad (y^- \neq 0) \qquad (8.18)$$

但若给定方程 $F(y, x_1, \cdots, x_m) = 0$ 不能直接解出 $y$，那又该如何呢？在这种情况下，若根据隐函数定理可知隐函数存在，则我们仍可不需解出 $y$ 而求得所需要的导数。为此，我们需要运用"隐函数法则"。通过该法则，我们可求出由给定方程所确定的每个隐函数的导数。隐函数法则的推导基于下述基本事实：(1) 若两个表达式恒等，则它们各自的全微分必定相等。[①](2) 对包含变量 $y, x_1, \cdots, x_m$ 的表达式微分，将会得到一个含有微分 $dy, dx_1, \cdots, dx_m$ 的表达式；(3) 若我们以 $dx_1$ 除 $dy$，并令所有其它的微分 $(dx_2, \cdots, dx_m)$ 为零，则商可看作偏导数 $\partial y/\partial x_1$。类似地，若我们以 $dx_2$ 除 $dy$，我们可以得到偏导数 $\partial y/\partial x_2$；如此等等。将上述事实应用于方程 $F(y, x_1, \cdots, x_m) = 0$（回忆一下，此方程在确定隐 208 函数的邻域内具有恒等性质），我们可以写成 $dF = 0$，或者

---

① 例如，取恒等式

$$x^2 - y^2 \equiv (x + y)(x - y)$$

由于 $x$ 和 $y$ 取任何值，方程两边均相等，所以此式为恒等式。对方程两边取全微分，我们有

$$\begin{aligned} d(\text{左边}) &= 2x\,dx - 2y\,dy \\ d(\text{右边}) &= (x - y)d(x + y) + (x + y)d(x - y) \\ &= (x - y)(dx + dy) + (x + y)(dx - dy) \\ &= 2xdx - 2ydy \end{aligned}$$

这两个结果实际上是相等的。然而，若这两个表达式不恒等，而只是对变量的某些特定的值相等，则全微分不相等。例如方程

$$x^2 - y^2 = x^2 + y^2 - 2$$

仅当 $y = \pm 1$ 时成立。两边的全微分为

$$\begin{aligned} d(\text{左边}) &= 2x\,dx - 2y\,dy \\ d(\text{右边}) &= 2x\,dx + 2y\,dy \end{aligned}$$

显然不等。特别要注意的是，它们在 $y = \pm 1$ 处也不等。

$$F_y\, dy + F_1\, dx_1 + \cdots + F_m dx_m = 0$$

假设仅允许 $y$ 和 $x_1$ 变化(即仅令 $dy$ 和 $dx_1$ 不为 0),则上面的方程简化为 $F_y\, dy + F_1\, dx_1 = 0$。以 $dx_1$ 除各项,并求解 $dy/dx_1$,得到:

$$\left.\frac{dy}{dx_1}\right|_{\text{其它变量不变}} \equiv \frac{\partial y}{\partial x_1} = -\frac{F_1}{F_y}$$

通过类似的方法,我们推导出隐函数 $f$ 的所有其它偏导数。这可以概括成一个一般法则——隐函数法则:给定 $F(y, x_1, \cdots, x_m) = 0$,若隐函数 $y = f(x_1, \cdots, x_m)$存在,则 $f$ 的偏导数为:

$$\frac{\partial y}{\partial x_i} = -\frac{F_i}{F_y} \ (i = 1, 2, \cdots, m) \tag{8.19}$$

一个简单的情形是给定方程是 $F(y, x) = 0$,由上述法则得出

$$\frac{dy}{dx} = -\frac{F_x}{F_y} \tag{8.19'}$$

这个法则表明,即使隐函数的具体形式为未知,我们仍可通过取函数 $F$ 的一对偏导数的比值的负值而求得隐函数的偏导数,而函数 $F$ 则是确定隐函数的已知方程 $F(y, x_1, \cdots, x_m) = 0$。要注意的是 $F_y$ 总是以比率的分母而出现。因此,不允许出现 $F_y = 0$ 的情况。由于隐函数定理设定在定义隐函数的某点附近 $F_y \neq 0$,因而在该点的相关邻域内分母为零的问题被自动排除了。

**例 1** 求由(8.14′)定义的隐函数的导数 $dy/dx$。因为 $F(y, x)$取 $y - 3x^4$ 的形式,由(8.19′),我们有:

$$\frac{dy}{dx} = -\frac{F_x}{F_y} = -\frac{-12x^3}{1} = 12x^3$$

在这个特定问题中,我们可以很容易解此方程得到 $y$,即 $y = 3x^4$。因此,用上述方法求得的导数的正确与否很容易得到验证。

**例2** 求由圆方程(8.15)所定义的隐函数的导数 $dy/dx$。这里我们有 $F(y,x)=x^2+y^2-9$,因而 $F_y=2y$, $F_x=2x$,由(8.19′),所求导数为



$$\frac{dy}{dx}=-\frac{2x}{2y}=-\frac{x}{y} \qquad (y\neq 0)^{①}$$

前面曾指出,由隐函数法则可求出由给定方程确定的每个隐函数的导数。下面,我们用(8.15′)中的两个函数和它们相应的导数(8.18)对此加以验证。若我们以 $y^+$ 替代用隐函数法则求得的结果 $dy/dx=-x/y$ 中的 $y$,我们便得到在(8.18)中所表示的结果 $dy^+/dx$;类似地,以 $y^-$ 替代 $y$,会得到(8.18)中的另一个导数。因而前面的结论得证。

**例3** 求由方程 $F(y,x,w)=y^3x^2+w^3+yxw-3=0$ 定义的任意隐函数的偏导数 $\partial y/\partial x$。此方程很难解出 $y$,但因 $F_y$, $F_x$, $F_w$ 明显连续,且 $F_y=3y^2x^2+xw$ 在满足给定方程的点,如(1,1,1)确实为非零,所以隐函数 $y=f(x,w)$ 至少在该点确实存在,因而讨论导数 $\partial y/\partial x$ 有意义。进而由(8.19),我们可立即得出

$$\frac{\partial y}{\partial x}=-\frac{F_x}{F_y}=-\frac{2y^3x+yw}{3y^2x^2+xw}$$

在点(1,1,1),该导数之值为 $-3/4$。

**例4** 假设方程 $F(Q,K,L)=0$ 隐含地定义了一个生产函数 $Q=f(K,L)$,让我们求出表示与函数 $F$ 相关的边际物质产品 $MPP_K$ 和 $MPP_L$ 的方法。因为边际产品仅为偏导数 $\partial Q/\partial K$ 和 $\partial Q/\partial L$,我们可应用隐函数法则并写出:

---

① $y\neq 0$ 的限制与前面对引出隐函数定理的讨论是非常一致的。

$$\text{MPP}_K \equiv \frac{\partial Q}{\partial K} = -\frac{F_K}{F_Q} \qquad \text{和} \qquad \text{MPP}_L \frac{\partial Q}{\partial L} = -\frac{F_L}{F_Q}$$

此外，我们还可从方程 $F(Q,K,L)=0$ 得到另一偏导数

$$\frac{\partial K}{\partial L} = -\frac{F_L}{F_K}$$

$\partial K/\partial L$ 的经济含义是什么呢？偏导数符号意味着另一变量 $Q$ 保持不变，由此可知此导数所描述的 $K$ 和 $L$ 的变化实质上是一种"补偿"变化，从而使产出 $Q$ 维持在某一特定水平不变，因而这种 210 变化属于沿着等产量曲线上的移动。实际上，导数 $\partial k/\partial L$ 是此等产量线斜率的度量，它在正常情况下为负。而 $\partial K/\partial L$ 的绝对值，则是两种投入的边际技术替代率的度量。

### 推广到联立方程组的情况

隐函数定理也可以更一般的形式出现，它讨论联立方程组

$$\begin{aligned}
F^1(y_1,\cdots,y_n;x_1,\cdots,x_m) &= 0 \\
F^2(y_1,\cdots,y_n;x_1,\cdots,x_m) &= 0 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
F^n(y_1,\cdots,y_n;x_1,\cdots,x_m) &= 0
\end{aligned} \tag{8.20}$$

此联立方程组定义了一组隐函数[①]

$$\begin{aligned}
y_1 &= f^1(x_1,\cdots,x_m) \\
y_2 &= f^2(x_1,\cdots,x_m) \\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned} \tag{8.21}$$

---

① 从另一方面考虑，这些条件可确保(8.20)中的 n 个方程原则上可解出 n 个变量 $y_1,\cdots,y_n$，尽管我们可能不能以显函数形式得到解(8.21)。

$$y_n = f^n(x_1, \cdots, x_m)$$

隐函数定理的一般化形式表达如下：

给定方程组(8.20)，若($a$)对所有的 $y$ 变量和 $x$ 变量，函数 $F^1$，…，$F^n$ 均具有连续偏导数，且若($b$)在点($y_{10}, \cdots, y_{n0}$；$x_{10}, \cdots, x_{m0}$)满足方程(8.20)，且下述雅可比行列式为非零：

$$|J| \equiv \left|\frac{\partial(F^1, \cdots, F^n)}{\partial(y_1, \cdots, y_n)}\right| \equiv \begin{vmatrix} \frac{\partial F^1}{\partial y_1} & \frac{\partial F^1}{\partial y_2} & \cdots & \frac{\partial F^1}{\partial y_n} \\ \frac{\partial F^2}{\partial y_1} & \frac{\partial F^2}{\partial y_2} & \cdots & \frac{\partial F^2}{\partial y_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial F^n}{\partial y_1} & \frac{\partial F^n}{\partial y_2} & \cdots & \frac{\partial F^n}{\partial y_n} \end{vmatrix} \neq 0$$

则存在一个($x_{10}, \cdots, x_{m0}$)的 $m$ 维邻域 $N$，在此邻域内，变量 $y_1, \cdots, y_n$ 是变量 $x_1, \cdots, x_m$ 的函数，其形式如(8.21)。这些隐函数满足

$$y_{10} = f^1(x_{10}, \cdots, x_{m0})$$
$$\cdots\cdots\cdots\cdots$$
$$y_{n0} = f^n(x_{10}, \cdots, x_{m0})$$

对邻域 $N$ 中的每个 $m$ 元数组($x_1, \cdots, x_m$)，它们也满足 211 (8.20)——因而在此邻域中使得(8.20)成为一组恒等式。而且，隐函数 $f^1, \cdots, f^n$ 连续，且对所有的 $x$ 变量具有连续偏导数。

同单一变量情况一样，可直接从(8.20)的 $n$ 个方程中解得隐函数的偏导数，而不需解出变量 $y$。利用在邻域 $N$ 中(8.20)为恒等式这一事实，我们可以对每个恒等式取全微分并写成 $dF^j = 0$

($j=1,2,\cdots,n$)。其结果是包括微分 $dy_1,\cdots,dy_n$ 和 $dx_1,\cdots,dx_m$ 的方程组。具体地,将 $dx_i$ 项移至等号右边后,我们有

$$
\begin{aligned}
&\frac{\partial F^1}{\partial y_1}dy_1+\frac{\partial F^1}{\partial y_2}dy_2+\cdots+\frac{\partial F^1}{\partial y_n}dy_n \\
&\qquad\qquad=-\left(\frac{\partial F^1}{\partial x_1}dx_1+\cdots+\frac{\partial F^1}{\partial x_m}dx_m\right) \\
&\frac{\partial F^2}{\partial y_1}dy_1+\frac{\partial F^2}{\partial y_2}dy_2+\cdots+\frac{\partial F^2}{\partial y_n}dy_n \\
&\qquad\qquad=-\left(\frac{\partial F^2}{\partial x_1}dx_1+\cdots+\frac{\partial F^2}{\partial x_m}dx_m\right) \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \qquad (8.22)\\
&\frac{\partial F^n}{\partial y_1}dy_1+\frac{\partial F^n}{\partial y_2}dy_2+\cdots+\frac{\partial F^n}{\partial y_n}dy_n \\
&\qquad\qquad=-\left(\frac{\partial F^n}{\partial x_1}dx_1+\cdots+\frac{\partial F^n}{\partial x_m}dx_m\right)
\end{aligned}
$$

因为在(8.22)中所有的偏导数在点($y_{10},\cdots,y_{n0};x_{10},\cdots,x_{m0}$)[在该点附近隐函数(8.21)有定义]处计算其值时,将取具体(常数)值,所以这里我们得到 $n$ 个线性方程组,在这个方程组中微分 $dy_1$(视为内生的)可用微分 $dx_i$ 来表示(可视 $dx_i$ 为外生的)。现在,除 $dx_1$ 外,我们令所有的 $dx_i$ 为零(即仅允许 $x_1$ 变化),则 $dx_2,\cdots,dx_m$ 各项将从方程组中消失。另外,假设我们以 $dx_1$ 除以余下各项,则出现表达式 $dy_1/dx_1,\cdots,dy_n/dx_1$。但是,我们可以将其理解成为(8.21)的偏导数,因为除 $x_1$ 外所有的变量 $x$ 均保持不变。因此,通过采取上述步骤,我们可以得到所要求的隐函数的偏导数。注意,实际上我们一举得到了全部 $n$ 个偏导数(在这里,它们是 $\partial y_1/\partial x_1,\cdots,\partial y_n/\partial x_1$)。

由上述步骤所得到的结果是下列线性方程组:

$$\frac{\partial F^1}{\partial y_1}\left(\frac{\partial y_1}{\partial x_1}\right)+\frac{\partial F^1}{\partial y_2}\left(\frac{\partial y_2}{\partial x_1}\right)+\cdots+\frac{\partial F^1}{\partial y_n}\left(\frac{\partial y_n}{\partial x_1}\right)=-\frac{\partial F^1}{\partial x_1}$$

$$\frac{\partial F^2}{\partial y_1}\left(\frac{\partial y_1}{\partial x_1}\right)+\frac{\partial F^2}{\partial y_2}\left(\frac{\partial y_2}{\partial x_1}\right)+\cdots+\frac{\partial F^2}{\partial y_n}\left(\frac{\partial y_n}{\partial x_1}\right)=-\frac{\partial F^2}{\partial x_1}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\frac{\partial F^n}{\partial y_1}\left(\frac{\partial y_1}{\partial x_1}\right)+\frac{\partial F^n}{\partial y_2}\left(\frac{\partial y_2}{\partial x_1}\right)+\cdots+\frac{\partial F^n}{\partial y_n}\left(\frac{\partial y_n}{\partial x_1}\right)=-\frac{\partial F^n}{\partial x_1} \tag{8.23}$$

这里为便于区分,我们为那些待求的导数加上括号以区别于现在看作常数的导数。用矩阵符号,此方程组可以表示成:

$$\begin{bmatrix} \frac{\partial F^1}{\partial y_1} & \frac{\partial F^1}{\partial y_2} & \cdots & \frac{\partial F^1}{\partial y_n} \\ \frac{\partial F^2}{\partial y_1} & \frac{\partial F^2}{\partial y_2} & \cdots & \frac{\partial F^2}{\partial y_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial F^n}{\partial y_1} & \frac{\partial F^n}{\partial y_2} & \cdots & \frac{\partial F^n}{\partial y_n} \end{bmatrix} \begin{bmatrix} \left(\frac{\partial y_1}{\partial x_1}\right) \\ \left(\frac{\partial y_2}{\partial x_1}\right) \\ \vdots \\ \left(\frac{\partial y_n}{\partial x_1}\right) \end{bmatrix} = \begin{bmatrix} -\frac{\partial F^1}{\partial x_1} \\ -\frac{\partial F^2}{\partial x_1} \\ \vdots \\ -\frac{\partial F^n}{\partial x_1} \end{bmatrix} \tag{8.23'}$$

因为(8.23′)中的系数矩阵只不过是特殊的雅可比行列式$|J|$,在隐函数定理的条件下,它不等于零。又因为方程组必是非齐次的(为什么?),所以方程组(8.23′)有唯一解。根据克莱姆法则,此解的解析表达式可表示成

$$\left(\frac{\partial y_j}{\partial x_1}\right)=\frac{|J_j|}{|J|} \quad (j=1,2,\cdots,n) \qquad [\text{见}(5.15)] \tag{8.24}$$

适当调整上述步骤,隐函数对其它变量 $x_2,\cdots,x_m$ 的偏导数也可以得到。

与单一方程相适应的隐函数法则(8.19)相似,在计算隐函数

$f^1,\cdots,f^n$ 的偏导数时，其步骤仅要求运用函数 $F$[在点$(y_{10},\cdots,y_{n0};x_{10},\cdots,x_{m0})$计算而来]的偏导数。因此，矩阵方程(8.23′)和其分析解(8.24)实际上表述了联立方程组的隐函数法则。

注意，要求$|J|\neq0$，就排除了(8.24)分母为零的情况，这与在隐函数法则(8.19)和(8.19′)中要求 $F_y\neq0$ 是一样的。同样，条件$|J|\neq0$ 在保证一般方程组(可能是非线性的)(8.20)具有唯一解(尽管是隐式解)方面所起的作用，与线性方程组 $Ax=d$ 中的非奇异条件$|A|\neq0$ 所起的作用是非常相似的。

213 **例 5** 将国民收入模型(7.17)重新写成下述形式

$$\begin{aligned} Y - C - I_0 - G_0 &= 0 \\ C - \alpha - \beta(Y - T) &= 0 \\ T - \gamma - \delta Y &= 0 \end{aligned} \tag{8.25}$$

若我们将内生变量$(Y,C,T)$取为$(y_1,y_2,y_3)$，将外生变量和参数$(I_0,G_0,\alpha,\beta,\gamma,\delta)$取为$(x_1,x_2,\cdots,x_6)$，则每个方程左边的表达式可以视为一个特定的 $F$ 函数，其形式为 $F^j(Y,C,T;I_0,G_0,\alpha,\beta,\gamma,\delta)$。因此，(8.25)是(8.20)的一个特例，其 $n=3,m=6$。因为函数 $F^1,F^2$ 和 $F^3$ 确实具有连续偏导数，且相应的雅可比行列式(仅包括内生变量的行列式)

$$|J| = \begin{vmatrix} \dfrac{\partial F^1}{\partial Y} & \dfrac{\partial F^1}{\partial C} & \dfrac{\partial F^1}{\partial T} \\ \dfrac{\partial F^2}{\partial Y} & \dfrac{\partial F^2}{\partial C} & \dfrac{\partial F^2}{\partial T} \\ \dfrac{\partial F^3}{\partial Y} & \dfrac{\partial F^3}{\partial C} & \dfrac{\partial F^3}{\partial T} \end{vmatrix} = \begin{vmatrix} 1 & -1 & 0 \\ -\beta & 1 & \beta \\ -\delta & 0 & 1 \end{vmatrix} = 1 - \beta + \beta\delta \tag{8.26}$$

总不为零($\beta$和$\delta$被限定为正分数),所以我们可以在满足(8.25)的任意一点附近把 $Y, C, T$ 看作$(I_0, G_0, \alpha, \beta, \gamma, \delta)$的隐函数。但满足(8.25)的点应是与 $\bar{Y}$、$\bar{C}$和 $\bar{T}$相关的均衡解,因此,根据隐函数定理我们可以写成

$$\bar{Y} = f^1(I_0, G_0, \alpha, \beta, \gamma, \delta)$$

$$\bar{C} = f^2(I_0, G_0, \alpha, \beta, \gamma, \delta)$$

$$\bar{T} = f^3(I_0, G_0, \alpha, \beta, \gamma, \delta)$$

此式表明内生变量的均衡值是外生变量及参数的隐函数。

隐函数的偏导数,如 $\partial\bar{Y}/\partial I_0$ 和 $\partial\bar{Y}/\partial G_0$ 等,实质上是比较静态导数。为求得这些导数,我们仅需计算模型在均衡状态时的 $F$ 函数的偏导数。进而,因 $n = 3$,所以通过一次运算可得到三个偏导数值。现在我们假设除 $G_0$ 外,所有其它外生变量和参数保持不变,则将此结果改写成(8.23′)的形式,我们可以写成方程

$$\begin{bmatrix} 1 & -1 & 0 \\ -\beta & 1 & \beta \\ -\delta & 0 & 1 \end{bmatrix} \begin{bmatrix} \partial\bar{Y}/\partial G_0 \\ \partial\bar{C}/\partial G_0 \\ \partial\bar{T}/\partial G_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

由此式可以计算出三个比较静态导数(均为对 $G_0$ 的导数)。例如,表示政府支出乘数的第一个导数为

214

$$\frac{\partial\bar{Y}}{\partial G_0} = \frac{\begin{vmatrix} 1 & -1 & 0 \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{vmatrix}}{|J|} = \frac{1}{1-\beta+\beta\delta} \quad [\text{由}(8.26)]$$

当然,这不过是前面在(7.19)中所得到的结果。但要注意,我们这

里所运用的方法只是对隐函数进行运算，完全不需解方程组(8.25)以求得 $\bar{Y}$，$\bar{C}$ 和 $\bar{T}$。正是由于这种方法的此类特征，才使得我们可以处理本质上难以有显式解的一般函数模型的比较静态分析问题。

## 练习 8.5

---

1　假设方程 $F(U, x_1, x_2, \cdots, x_n) = 0$ 隐式定义了一个效用函数 $U = f(x_1, x_2, \cdots, x_n)$：

(a) 求 $\partial u / \partial x_2$，$\partial u / \partial x_n$，$\partial x_3 / \partial x_2$，$\partial x_4 / \partial x_n$ 的表达式。

(b) 分别解释其经济含义。

2　给定如下方程 $F(y, x) = 0$：

(a) $x^3 - 2x^2 y + 3xy^2 - 22 = 0$

(b) $2x^2 + 4xy - y^4 + 67 = 0$

在点($y = 3, x = 1$)附近能确定一个隐函数 $y = f(x)$ 吗？若你的答案是肯定的，用隐函数法则求 $dy/dx$，并在点($y = 3, x = 1$)处计算其值。

3　已知 $x^2 + 3xy + 2yz + y^2 + z^2 - 11 = 0$，在点($x = 1, y = 2, z = 0$)附近能否定义隐函数 $y = f(x, z)$？若能，运用隐函数法则求 $\partial z / \partial x$，$\partial z / \partial y$，并计算其在该点的值。

4　在原点附近的邻域内，考察方程 $F(y, x) = (x - y)^3 = 0$，证明隐函数定理所提到的条件并非必要条件。

5　如果方程 $F(x, y, z) = 0$ 将三个变量中的每个变量定义为另两个变量的隐函数，且如果所要求的导数均存在，求下式的值：

$$\frac{\partial z}{\partial x} \frac{\partial x}{\partial y} \frac{\partial y}{\partial z}$$

6 验证课文中的结论“方程组(8.23′)必定为非齐次的”是正确的。

7　由国民收入模型(8.25)，运用隐函数法则求非所得税乘数和所得税率乘数。与(7.20)和(7.21)相比较，验证你的结果。

---

## 8.6 一般函数模型的比较静态学



当我们在第七章初次考察比较静态分析问题时，我们讨论了这样一种情况：模型内生变量的均衡值可由外生变量和参数显式表示。在那里，简单的偏微分技术已足以解决全部问题。但是，当模型含有以一般形式表示的函数时，由于难以得到显示解，偏微分技术已难以适应。因此必须使用像全微分、全导数，以及隐函数定理、隐函数法则等新的方法。我们首先用市场模型，然后再运用国民收入模型来介绍这些方法。

### 市场模型

考察一个单一商品市场，其中需求量 $Q_d$ 不仅是价格 $P$，而且是外生确定的收入 $Y_0$ 的函数，但供给量 $Q_s$ 则仅是价格的函数。如果这些函数并未以具体形式给出，则我们可以将这个模型一般地写成：

$$\begin{aligned} Q_d &= Q_s \\ Q_d &= D(P, Y_0) \quad (\partial D/\partial P < 0; \partial D/\partial Y_0 > 0) \\ Q_s &= S(P) \qquad\quad (dS/dP > 0) \end{aligned} \tag{8.27}$$

假设函数 $D$ 和 $S$ 均拥有连续偏导数，或者换句话说，均具有平滑的曲线；而且，为了保证其经济意义，我们对这些导数的符号施加明确的限制。尽管供给函数可以是线性的，也可以是非线性的，但限制条件 $dS/dP>0$，规定了供给函数是单调递增函数。类似地，对需求函数的两个偏导数符号的限制可以表明它是价格的减函数，收入的增函数。这些限制可以把我们的分析限定在我们希望遇到的“正常”情况。

在描绘通常的二维需求曲线时,收入水平被假定为固定不变。当收入变化时,由于会导致需求曲线移动而破坏给定的均衡。类似地,在(8.27)中,$Y_0$ 可通过需求函数导致非均衡变化。这里,$Y_0$ 是唯一的外生变量或参数,所以此模型的比较静态分析就只关注 $Y_0$ 的变化如何影响模型的均衡状态。

市场的均衡状态由均衡条件 $Q_d = Q_s$ 所确定。通过替代和重排,均衡条件可以表示成:

$$D(P, Y_0) - S(P) = 0 \tag{8.28}$$

216 尽管不能解此方程求出均衡价格 $\bar{P}$,但我们仍假设确实存在静态均衡——否则即便提出比较静态分析问题都没有意义。根据我们处理具体函数模型的经验,我们可以预期 $\bar{P}$ 是外生变量 $Y_0$ 的函数:

$$\bar{P} = \bar{P}(Y_0) \tag{8.29}$$

现在我们借助于隐函数定理,对这种预期提供严格的依据。因为(8.28)的形式是 $F(P, Y_0) = 0$,满足隐函数定理的条件将会保证在满足(8.28)的某一点的邻域内,即在均衡(初始或旧的)解的邻域内,每一个 $Y_0$ 值都得到一个唯一的 $\bar{P}$ 值。在此情况下,我们实际上可以写出隐函数 $\bar{P} = \bar{P}(Y_0)$,并讨论其导数 $d\bar{P}/dY_0$——我们知道它是存在的,它正是我们所要求的比较静态导数。现在我们来检验那些条件。首先,函数 $F(\bar{P}, Y_0)$ 确实具有连续导数,因为根据假设,函数和的两个部分 $D(P, Y_0)$ 和 $S(P)$ 均具有连续导数;其次,函数 $F$ 对 $P$ 的偏导数,即 $F_P = \partial D/\partial P - \partial S/\partial P$ 为负,因此无论在何处计算均不等于零。因此,可应用隐函数定理,且(8.29)确实成立。

基于同样的定理,均衡条件(8.28)在均衡解的某一邻域内可

视作恒等式。这样,我们可以把均衡等式写成

$$D(\bar{P}, Y_0) - S(\bar{P}) \equiv 0 \tag{8.30}$$

则只需直接应用隐函数法则便可得到比较静态导数 $d\bar{P}/d\bar{Y}_0$。为便于识别,以后我们将导数 $d\bar{P}/dY_0$ 加上括号以区别于一般的导数。这些导数只是模型特征的一部分。比较静态导数的结果是

$$\left(\frac{d\bar{P}}{dY_0}\right) = -\frac{\partial F/\partial Y_0}{\partial F/\partial \bar{P}} = -\frac{\partial D/\partial Y_0}{\partial D/\partial \bar{P} - dS/d\bar{P}} > 0 \tag{8.31}$$

在此结果中,表达式 $\partial D/\partial \bar{P}$ 是导数 $\partial D/\partial P$ 在初始均衡点 $P = \bar{P}$ 处计算的值。对 $dS/d\bar{P}$ 也可以作类似的解释。事实上,$\partial D/\partial Y_0$ 也必须在均衡点计算。由于(8.27)中符号的设定,$(d\bar{P}/dY_0)$ 恒为正,因此我们的定量结论是:收入水平的提高(下降)将会导致均衡价格的提高(下降)。如果供给函数和需求函数在初始均衡的导数值为已知,则(8.31)当然也会给出定量的结论。

上述讨论涉及到 $Y_0$ 变化对 $\bar{P}$ 的影响。那么,能否发现 $Y_0$ 变化对均衡数量 $\bar{Q}(=\bar{Q}_d=\bar{Q}_s)$ 的影响呢?答案是肯定的。因为在均衡状态,我们有 $\bar{Q} = S\bar{P})$,又因为 $\bar{P} = \bar{P}(Y_0)$,我们可应用链式法则得到导数 217

$$\left(\frac{d\bar{Q}}{dY_0}\right) = \frac{dS}{d\bar{P}}\left(\frac{d\bar{P}}{dY_0}\right) > 0 \quad \left[\text{因为}\frac{dS}{d\bar{P}} > 0\right] \tag{8.32}$$

因此,在此模型中,均衡数量也与 $Y_0$ 正相关。而且,如果各导数在均衡时的取值已知,(8.32)也会给出定量的结论。

(8.31)和(8.32)包括了市场模型中所有的比较静态分析的内容,这些结果并不出人意料。事实上,它们只不过传递了这样一个命题:需求曲线向上移动将会导致更高的均衡价格和更大的均衡

数量。好像只用简单的图解分析就可以得到同样的命题，这似乎有道理，但人们不应忽略我们这里所用的分析方法具有更普遍的一般性。再重复一遍：图形分析就其本质而言仅局限于一组具体的曲线（即一组特定函数的几何表示），因此严格地说，其结论也仅与这组曲线相联系，仅适应于这组曲线。与此形成鲜明对照的是，(8.27)式虽然简单，却包含了斜率为负的需求曲线和斜率为正的供给曲线所有可能组合的全部集合，这样它也就更为一般化。此外，这里采用的分析方法还可处理图形分析方法难以解决的远为复杂的问题。

## 联立方程法

模型(8.27)的分析是以一个单一方程即(8.30)为基础完成的。由于一个方程只能包含一个内生变量，所以包含了 $\bar{P}$ 则意味着排除了 $\bar{Q}$。因而我们不得不首先求出$(d\bar{P}/dY_0)$，然后在下一步再导出$(d\bar{Q}/dY_0)$。现在我们来介绍如何同时研究 $\bar{P}$ 和 $\bar{Q}$。因为有两个内生变量，相应地我们要建立由两个方程组成的方程组。首先，令(8.27)中的 $Q = Q_d = Q_s$，并重排，我们可将市场模型表示成

$$\begin{aligned} F^1(P,Q;Y_0) &= D(P,Y_0) - Q = 0 \\ F^2(P,Q;Y_0) &= S(P) - Q = 0 \end{aligned} \tag{8.33}$$

此式与(8.20)的形式一致，其中 $n=2, m=1$。再一次检验隐函数定理的条件是有意义的。首先，因需求与供给函数均假定有连续偏导数，所以函数 $F^1$ 与 $F^2$ 必定也具有连续偏导数。其次，内生变量雅可比行列式（包含 $P$ 和 $Q$ 的雅可比行列式）确实不为零，不

管在哪一点计算其值。因为：

$$|J| = \begin{vmatrix} \frac{\partial F^1}{\partial P} & \frac{\partial F^1}{\partial Q} \\ \frac{\partial F^2}{\partial P} & \frac{\partial F^2}{\partial Q} \end{vmatrix} = \begin{vmatrix} \frac{\partial D}{\partial P} & -1 \\ \frac{dS}{dP} & -1 \end{vmatrix} = \frac{dS}{dP} - \frac{\partial D}{\partial P} > 0 \quad (8.34)$$

218

因此，如果均衡解存在（我们必须作这样的假定才能使我们对比较静态学的讨论有意义），根据隐函数定理我们可以写出隐函数：

$$\bar{P} = \bar{P}(Y_0) \quad 和 \quad \bar{Q} = \bar{Q}(Y_0) \quad (8.35)$$

尽管我们不能解出 $\bar{P}$ 和 $\bar{Q}$。我们知道，这些函数具有连续导数，而且在均衡状态的某一邻域内，(8.33)是一对恒等式，所以我们也可以写成

$$\begin{aligned} D(\bar{P}, Y_0) - \bar{Q} &\equiv 0 \\ S(\bar{P}) - \bar{Q} &\equiv 0 \end{aligned} \quad (8.36)$$

由此，$(d\bar{P}/DY_0)$和$(d\bar{Q}/dY_0)$可同时得到。

微分 $d\bar{P}$，$\mathrm{d}\bar{Q}$和 $dY_0$ 是这两个导数的组成部分。为得到这些微分表达式，我们对(8.36)中的每个恒等式依次进行微分，重排后得到关于 $d\bar{P}$和 $d\bar{Q}$的线性方程组

$$\frac{\partial D}{\partial \bar{P}} \mathrm{d}\bar{P} - d\bar{Q} = -\frac{\partial D}{\partial Y_0} dY_0$$

$$\frac{dS}{d\bar{P}} \mathrm{d}\bar{P} - d\bar{Q} = 0$$

因为变量 $d\bar{P}$ 和 $d\bar{Q}$ 均为一次幂，所有在初始均衡状态时计算其值的系数导数均代表常数，外生变量的任意非零变化 $dY_0$ 也代表

常数，所以，此方程组是线性方程组。以 $dY_0$ 通除各项，并将两个微分的商视为导数，我们得到矩阵方程①

$$\begin{bmatrix} \frac{\partial D}{\partial \bar{P}} & -1 \\ \frac{dS}{d\bar{P}} & -1 \end{bmatrix} \begin{bmatrix} \left(\frac{d\bar{P}}{dY_0}\right) \\ \left(\frac{d\bar{Q}}{dY_0}\right) \end{bmatrix} = \begin{bmatrix} -\frac{\partial D}{\partial Y_0} \\ 0 \end{bmatrix}$$

219 由克莱姆法则，并利用(8.34)，我们得到解为：

$$\left(\frac{d\bar{P}}{dY_0}\right) = \frac{\begin{vmatrix} -\frac{\partial D}{\partial Y_0} & -1 \\ 0 & -1 \end{vmatrix}}{|J|} = \frac{\frac{\partial D}{\partial Y_0}}{|J|}$$

$$\left(\frac{d\bar{Q}}{dY_0}\right) = \frac{\begin{vmatrix} \frac{\partial D}{\partial \bar{P}} & -\frac{\partial D}{\partial Y_0} \\ \frac{dS}{d\bar{P}} & 0 \end{vmatrix}}{|J|} = \frac{\frac{dS}{d\bar{P}} \frac{\partial D}{\partial Y_0}}{|J|} \tag{8.37}$$

其中需求与供给函数(包括那些在雅可比行列式中的供求函数)的所有导数均在初始均衡处计算其值。读者可以验证，刚才所得到的结果与前面通过单一方程方法所得到的结果(8.31)和(8.32)是一致的。

## 全导数方法的运用

上面所介绍的单一方程法和联立方程组法均具有一个共同的特点：我们都对均衡恒等式两边取全微分，然后令其相等。然而，

① 不取全微分并以 $dY_0$ 通除，采用隐函数法则(8.23)仍可得到同样的矩阵方程。

不取全微分，仍可能对特定的外生变量或参数取全导数，并令其结果相等。

例如，在单一方程法中，均衡恒等式为

$$D(\bar{P}, Y_0) - S(\bar{P}) \equiv 0 \qquad [由 (8.30)]$$

其中 $\bar{P} = \bar{P}(Y_0)$ [由 (8.29)]

在均衡恒等式两边对 $Y_0$ 取全导数(这不仅考虑到了 $Y_0$ 变化的直接影响，而且也考虑到了间接影响)，因而得到方程

$$\frac{\partial D}{\partial \bar{P}}\left(\frac{d\bar{P}}{dY_0}\right) + \frac{\partial D}{\partial Y_0} - \frac{dS}{d\bar{P}}\left(\frac{d\bar{P}}{dY_0}\right) = 0$$

($Y_0$ 对 $D$ 的间接影响)　($Y_0$ 对 $D$ 的直接影响)　($Y_0$ 对 $S$ 的间接影响)

将此式解出$(d\bar{P}/dY_0)$，其结果与(8.31)是一致的。

220

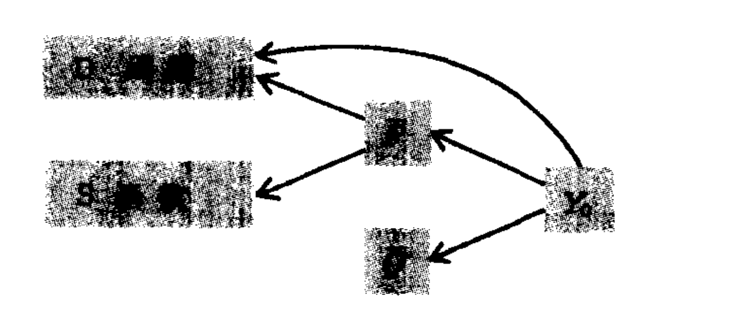


图 8.7

而在联立方程法中，有两个均衡恒等式

$$D(\bar{P}, Y_0) - \bar{P} \equiv 0$$

$$S(\bar{P}) - \bar{Q} \equiv 0 \qquad [由(8.36)]$$

其中 $\bar{P} = \bar{P}(Y_0)$　$\bar{Q} = \bar{Q}(Y_0)$ [由(8.35)]

显然，$Y_0$ 变化的各种影响很难识别，但借助于图 8.7 的通道图，影响的模式就变得非常明了了。比如，由此通道图可知，将函数 $D$

对 $Y_0$ 微分时，我们必须考虑到 $Y_0$ 通过 $\bar{P}$ 对 $D$ 的间接影响，以及 $Y_0$ 的直接影响(曲线箭头)。另一方面，在将函数 $S$ 对 $Y_0$ 微分时，则只需考虑通过 $\bar{P}$ 对 $S$ 的间接影响。所以经重新整理后，两个恒等式对 $Y_0$ 微分的结果是下面两个方程

$$\frac{\partial D}{\partial \bar{P}}\left(\frac{d\bar{P}}{\mathrm{d}Y_0}\right)-\left(\frac{d\bar{Q}}{\mathrm{d}Y_0}\right)=-\frac{\partial D}{\partial Y_0}$$

$$\frac{dS}{d\bar{P}}\left(\frac{d\bar{P}}{\mathrm{d}Y_0}\right)-\left(\frac{d\bar{Q}}{\mathrm{d}Y_0}\right)=0$$

当然，它们与通过全微分法得到的方程是一致的，而且由它们可再次导出(8.37)中的比较静态导数。

## 国民收入模型

现在将刚才介绍的方法应用于以一般函数表示的国民收入模型。为增加一些变化，这次我们在模型中减去政府支出和税收，而加上对外贸易。另外，我们还把货币市场与商品市场置于一起。

更具体地，假定商品市场的特征由以下四个函数规定：

1. 投资支出 $I$ 是利息率 $i$ 的减函数：

$$I = I(i) \qquad (I' < 0)$$

其中 $I' \equiv dI/di$，是投资函数的导数。

221

2. 储蓄 $S$ 是国民收入 $Y$ 以及利息率 $i$ 的增函数，边际储蓄倾向为正分数：

$$S = S(Y, i) \qquad (0 < S_Y < 1; S_i > 0)$$

其中 $S_Y = \partial S/\partial Y$ (边际储蓄倾向)和 $S_i \equiv \partial S/\partial i$ 均为偏导数。

3. 进口支出 $M$ 是国民收入的函数，边际进口倾向也是正分数：

$M = M(Y) \quad (0 < M' < 1)$

4. 出口水平 $X$ 由外生因素决定：

$X = X_0$

在货币市场中，我们有如下两个函数：

5. 货币需求量 $M_d$ 是国民收入的增函数（交易需求），但是利息率的减函数（投机需求）：

$M_d = L(Y, i) \quad (L_Y > 0; L_i < 0)$

这里之所以使用函数符号 $L$，是因为货币需求函数习惯上被称作流动偏好函数（Liquidity function）。$M_d$ 代表货币需求，应注意避免与表示进口的符号 $M$ 相混淆。

6. 货币供给作为一个货币政策问题，由外生因素确定：

$M_s = M_{s0}$

注意，$I$、$S$、$M$ 和 $X$ 同 $Y$ 一样，均是流量概念，是在一定时间内度量的；而 $M_d$ 和 $M_s$ 则是存量概念，它们表示在某一特定时点存在的量。但无论是存量还是流量，上述函数均被假定具有连续导数。

此模型要达到均衡需要同时满足商品市场均衡条件（注入 = 漏出，或者 $I + X = S + M$）以及货币市场的均衡条件（货币需求 = 货币供给，或 $M_d = M_s$）。在上述一般函数的基础上，均衡状态可通过下述两个条件来表示

$$\begin{aligned} I(i) + X_0 &= S(Y, i) + M(Y) \\ L(Y, i) &= M_{s0} \end{aligned} \tag{8.38}$$

由于符号 $I$、$S$、$M$ 和 $L$ 可视为函数符号，实际上我们只有两

个内生变量收入 $Y$ 和利率 $i$，以及两个外生变量，出口 $X_0$（由外国

222 决定）和 $M_{s0}$（由货币当局决定）。因此(8.38)可以(8.20)形式表示，其中 $n=m=2$：

$$
\begin{aligned}
F^1(Y,i;X_0,M_{s0}) &= I(i)+X_0-S(Y,i)-M(Y)=0\\
F^2(Y,i;X_0,M_{s0}) &= L(Y,i)-M_{s0}=0
\end{aligned}
\tag{8.38'}
$$

此方程组满足隐函数定理的条件，因为(1) $F^1$ 和 $F^2$ 具有连续导数（根据假设，所有函数的导数连续）；(2) 在初始均衡处（假设它存在）和其它地方，内生变量的雅可比行列式不为零：

$$
|J| = \begin{vmatrix} \partial F^1/\partial Y & \partial F^1/\partial i \\ \partial F^2/\partial Y & \partial F^2/\partial i \end{vmatrix} = \begin{vmatrix} -S_Y-M' & I'-S_i \\ L_Y & L_i \end{vmatrix}
$$

$$
= -L_i(S_Y+M')-L_Y(I'-S_i)>0 \tag{8.39}
$$

因此，尽管我们不能明确解出 $\bar{Y}$ 和 $\bar{i}$，但可以写出隐函数

$$
\bar{Y}=\bar{Y}(X_0,M_{s0}) \text{ 和 } \bar{i}=\bar{i}(X_0,M_{s0}) \tag{8.40}
$$

进而，我们把(8.38′)看作均衡点某一邻域的一对恒等式，所以我们可以写出

$$
\begin{aligned}
I(\bar{i})+X_0-S(\bar{Y},\bar{i})-M(\bar{Y}) &\equiv 0\\
L(\bar{Y},\bar{i})-M_{s0} &\equiv 0
\end{aligned}
\tag{8.41}
$$

由这些均衡恒等式，可以得到四个比较静态导数，两个与 $X_0$ 有关，另两个与 $M_{s0}$ 相关。但这里我们只推导出前两个导数，后两个留给读者作练习。

因此，在对(8.41)的每个恒等式取全微分以后，令 $dM_{s0}=0$，从而使 $dx_0$ 成为唯一的不均衡因子。其次，以 $dx_0$ 通除，并把两

个微分的商视为偏导数(因为另外一个外生变量 $M_{s0}$保持不变),得到矩阵方程:

$$\begin{bmatrix} -S_Y - M' & I' - S_i \\ L_Y & L_i \end{bmatrix} \begin{bmatrix} (\partial \bar{Y} / \partial X_0) \\ (\partial \bar{i} / \partial X_0) \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \qquad (8.42)$$

由克莱姆法则并运用(8.39),得到解

$$\left(\frac{\partial \bar{Y}}{\partial X_0}\right) = \frac{\begin{vmatrix} -1 & I' - S_i \\ 0 & L_i \end{vmatrix}}{|J|} = \frac{-L_i}{|J|} > 0$$

$$\left(\frac{\partial \bar{i}}{\partial X_0}\right) = \frac{\begin{vmatrix} -S_Y - M' & -1 \\ L_Y & 0 \end{vmatrix}}{|J|} = \frac{L_Y}{|J|} > 0 \qquad (8.43)$$

其中等号右边所有的导数(包括那些出现在雅可比行列式中的导 223 数)都在初始均衡,即 $Y = \bar{Y}, i = \bar{i}$ 处计算其值。当这些导数的具体值为已知时,(8.43)产生关于出口变化的影响的定量结论。但当这些值为未知时,只能得到定性结论:在此模型中 $\bar{Y}$ 和 $\bar{i}$ 将随出口增加而增加。

像在市场模型中一样,如果不使用全微分法,我们还有一个选择是将(8.41)中的均衡恒等式对所研究的特定的外生变量,即 $X_0$ 取**全导数**。当然,我们在取全导数时必须记住隐式解(8.40)。如在(8.41)和(8.40)中给出的那样,$X_0$ 影响模型不同部分的各种方式通过通道图 8.8 概括出来。特别应注意的是,对 $X_0$ 微分储蓄函数或流动偏好函数时,我们必须考虑到通过 $\bar{i}$ 和 $\bar{Y}$ 两方面的间接影响。借助于通道图,我们可以对 $X_0$ 微分均衡恒等式,以得到下面两个方程:

$$I'\left(\frac{\partial \bar{i}}{\partial X_0}\right)+1-S_Y\left(\frac{\partial \bar{Y}}{\partial X_0}\right)-S_i\left(\frac{\partial \bar{i}}{\partial X_0}\right)-M'\left(\frac{\partial \bar{Y}}{\partial X_0}\right)=0$$

$$L_Y\left(\frac{\partial \bar{Y}}{\partial X_0}\right)+L_i\left(\frac{\partial \bar{i}}{\partial X_0}\right)=0$$

因为另一个外生变量 $M_{s0}$ 保持不变，这两个方程左侧表示对应的两个均衡恒等式中左边表达式的偏全导数。然而，作为隐函数(8.40)的导数，比较静态导数$(\partial \bar{Y}/\partial X_0)$和$(\partial \bar{i}/\partial X_0)$仅仅是简单

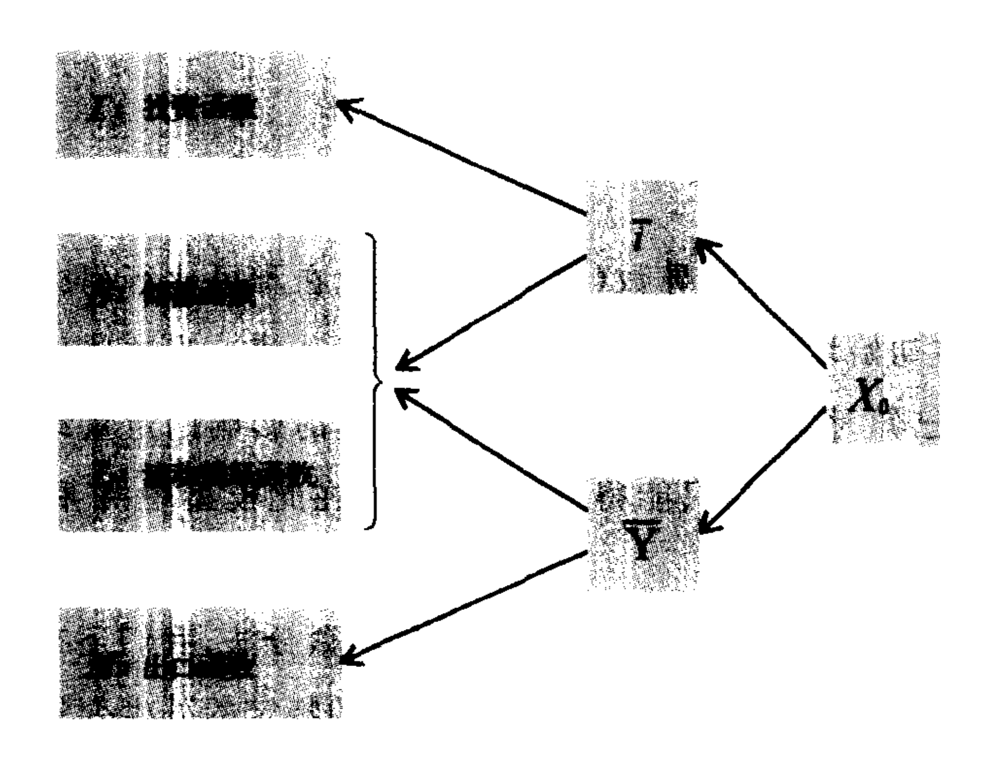



图 8.8

224 的偏导数。作适当的压缩，这两个方程可以压缩成(8.42)。所以，全微分法与全导数法得到了一致的结果。

读者应注意到，$(\partial \bar{Y}/\partial X_0)$实质上是出口乘数。因为出口引致的均衡收入的增加，通过进口函数 $M=M(Y)$使进口也增加；我们可再应用链式法则求得（辅助）比较静态导数：

$$\left(\frac{\partial \bar{M}}{\partial X_0}\right) = M'\left(\frac{\partial \bar{Y}}{\partial X_0}\right) = \frac{-M'L_i}{|J|}$$

因为 $M'>0$，所以上述导数符号为正。用完全类似的步骤也可求出其他辅助比较静态导数，例如$(\partial \bar{I}/\partial X_0)$，$(\partial \bar{S}/\partial X_0)$等。

## 步骤总结

在分析一般函数市场模型和国民收入模型时，不可能得到内生变量显式解的值。相反，我们依靠隐函数定理写出下面的隐函数：

$$\bar{P} = \bar{P}(Y_0) \quad 和 \quad \bar{i} = \bar{i}\,(X_0, M_{s0})$$

我们下一步求出诸如$(d\bar{P}/dY_0)$和$(\partial \bar{i}/\partial X_0)$等的比较静态导数，然后基于根据隐函数定理得知的事实解释函数 $\bar{P}$ 和 $\bar{i}$ 具有连续导数的意义。

为便于应用这个定理，我们使以(8.16)或(8.20)形式写出模型的均衡条件成为标准的作法，然后检验：(1) 函数 $F$ 是否具有连续导数；(2) $F_y$ 的值或者内生变量雅可比行列式(取决于具体情况)在模型初始均衡处是否为非零。然而，只要模型中的单个函数具有连续导数——在一般函数模型中经常采用此假设是很自然的——上述第一个条件便自动得到满足。因此，实际上只需检验 $F_y$ 的值或者内生变量雅可比行列式的值。如果在均衡处不为零，则我们便可以立刻求出比较静态导数。

要达到此目的，隐函数法则是有用的。在单一方程情况下，仅需令内生变量等于均衡条件下的均衡值，即令 $P=\bar{P}$，然后应用(8.19)中所述的法则于所得到的均衡恒等式。对于联立方程组的情况，我们也必须首先令所有内生变量分别等于均衡状态时的均衡值；然后我们再对所得到的均衡恒等式应用(8.24)所述的隐函

225 数法则，或者按下面介绍的几个步骤进行：

1. 对每个均衡恒等式依次取全微分。
2. 选择一个且仅选择一个外生变量，比如 $X_0$ 作为唯一的非均衡因子，并令所有其它外生变量的微分等于零。然后以 $dX_0$ 除以每个恒等式中余下各项，并将两个微分的商视为比较静态导数——若模型包含两个或两个以上的外生变量，应视作一个偏导数。①
3. 解所得到的方程组，求出比较静态导数，并解释其经济意义。在这一步，如果使用克莱姆法则，则可以利用这一事实：即前面在检验 $|J| \neq 0$ 条件时，我们已经计算出现在要解的方程组的系数矩阵行列式的值。
4. 若有其它非均衡因子（其它外生变量），其分析可重复步骤 2 和步骤 3。尽管在新的方程组中会出现一组不同的比较静态导数，系数矩阵却会同以前一样，所以可再次运用已知的 $|J|$ 值。

如果给定的模型具有 $m$ 个外生变量，要求得所有的比较静态导数，则需应用上述步骤 $m$ 次。

## 练习 8.6

1 令国民收入的均衡条件是

① 若不采取步骤 1 和步骤 2，我们也可以采用全导数的方法，将每个均衡恒等式的两边对选定的外生变量取全微分。若这样做，通道图就是很有益的了。

$$S(Y)+T(Y)=I(Y)+G_0(S',\ T',\ I'>0;\ S'+T'>I')$$

其中 $S,Y,T,I$ 和 $G$ 分别代表储蓄、国民收入、税收、投资和政府支出。所有的导数均连续。

($a$) 解释导数 $S',T'$ 和 $I'$ 的经济含义。

($b$) 检验隐函数定理的条件是否满足。若满足,写出均衡恒等式。

($c$) 求$(d\bar{Y}/dG_0)$并讨论其经济意义。

2 令某一商品的需求和供给函数为

$$Q_d=D(P,Y_0) \qquad (D_p<0;\ D_{Y_0}>0)$$
$$Q_s=S(P,T_0) \qquad (S_p>0;\ S_{T_0}<0)$$

其中 $Y_0$ 为收入,$T_0$ 为对此商品征的税。所有导数均连续。 226

($a$) 写出单一方程的均衡条件。

($b$) 检验能否应用隐函数定理。若能,写出均衡恒等式。

($c$) 求$(\partial\bar{P}/\partial Y_0)$和$(\partial\bar{P}/\partial T_0)$,并讨论其经济意义。

($d$) 运用与(8.32)类似的步骤,由供给函数求$(\partial\bar{Q}/\partial Y_0)$;由需求函数求$(\partial\bar{Q}/\partial T_0)$。[为什么不由需求函数求$(\partial\bar{Q}/\partial Y_0)$,由供给函数求$(\partial\bar{Q}/\partial T_0)$呢?]

3 以联立方程法解上题。

4 假设某一商品的需求函数和供给函数为:

$$Q_d=D(P,t_0)\ \left(\frac{\partial D}{\partial P}<0;\ \frac{\partial D}{\partial t_0}>0\right) \text{和}\ Q_s=Q_{s0}$$

其中 $t_0$ 是消费者对此商品的偏好,而且两个偏导数均连续。

($a$) 写出单一方程的均衡条件。

($b$) 能够运用隐函数定理吗?

($c$) 均衡价格将如何随着消费者偏好的变化而变化?

5 由(8.38)的国民收入模型,求$(\partial\bar{Y}/\partial M_{s0})$和$(\partial\bar{i}/\partial M_{s0})$,并解释其经济意义。求上述导数值既要全微分法,又要用全导数法,并证明两种方法所得的结果相同。

6 考察下列国民收入模型(忽略税收):

$$Y-C(Y)-I(i)-G_0=0 \qquad (0<C'<1;I'<0)$$
$$kY+L(i)-M_{s0}=0 \qquad (k=\text{正常数};L'<0)$$

($a$) 第一个方程具有均衡条件的性质吗?

($b$) 在此模型中,货币需求总量是多少?

(c) 当货币供给变化(货币政策)和政府支出变化(财政政策)时,分析模型的比较静态结果。

---

## 8.7 比较静态学的局限性

由于在经济学中,我们常对发现一个参数的非均衡变化如何影响模型的均衡状态感兴趣,所以比较静态分析是一个有价值的研究领域。但是,认识到比较静态学的这一性质是非常重要的:即比较静态学忽略了从旧均衡向新均衡的调整过程,同时它也忽略了调整过程中的时间因素。结果,它必然忽略了由于模型内在的不稳定性,从而导致新均衡难以达到的可能性。对调整过程的研究属于**动态经济学**的范畴。当我们进入动态经济学的范畴时,我们将会特别关注变量随时间变化的方式,以及均衡的稳定性问题。227

但是我们必须暂缓对动态经济学这一重要问题的研究。在本书的下一部分,我们将研究最优化问题,它也是一种具有比较静态含义的非常重要的一类特殊均衡分析。

# 第四篇

# 最优化问题

231

# 第9章　最优化:一类特殊的均衡分析

在第三章首次引入均衡概念时,我们在目标均衡与非目标均衡之间作了大致的区分。以我们研究的市场模型和国民收入模型为例,非目标均衡是指模型中的某些相反的力量——比如市场模型中的供给和需求以及国民收入模型中的注入与漏出——恰好处于彼此相等,相互平衡的状态,因而排除了进一步变化的趋势。这种均衡的实现是这些力量非人为平衡的结果,不需要有关参与人有意识地努力以实现特定目标。诚然,需求力量背后的居民户和供给力量背后的厂商在给定环境下都会主动谋求一种最佳的地位,但仅就市场本身而言,没有人会追求任何一种特定的均衡价格和均衡数量,除非政府恰好试图固定价格。与此相似,在国民收入决定中,漏出与注入的平衡会导致均衡状态,不需任何旨在实现特定目标(比如通过货币政策或财政政策以调整非意愿的收入水平等)的人为努力。

232　但在本书的这一部分,我们将着意研究目标均衡。所谓目标均衡是指给定经济单位,如居民户、厂商或者整个经济等的最优状态;而且这些经济单位主动谋求均衡的实现。因此,在这里,而且仅仅是在这里,我们前面提醒大家注意的"均衡并不意味着人为需要",在这里就变得无关紧要,甚至不恰当了。在本书这一部分中,我们将主要介绍确定最优化的古典方法——微积分法。较为现代

的方法，如数学规划等，将在后面讨论。

## 9.1 最优值与极值

经济学基本上是一门关于选择的科学。当要实现一个特定的经济目标，如要实现一个特定水平的产出，通常有许多可供选择的方式。但是在这诸多选择中，按照某一标准，会有一种方式会比其它方式更好；根据所规定的标准，选择最适宜的方式，这就是最优化问题的实质。

在经济学中，最常见的选择标准是最大化目标（如厂商利润最大化、消费者效用最大化、厂商或一国经济增长率最大化等）或最小化目标（如在给定产出下使成本最低等）。在经济学上，我们可以把最大化问题和最小化问题归为最优化问题这一类，表示“寻求最优”。但从纯数学角度看，“极大值”和“极小值”这两个术语并无最优化的含义。因此，作为数学概念，极大值和极小值这两个术语实际上表示极值。

系统地阐述一个最优化问题，首先要确定目标函数，其中因变量表示最大化或最小化的对象；而自变量则表示这样一组对象，其大小由所涉及的经济单位出于最优化的考虑而进行选择。因此，我们将这些自变量称作选择变量[①]。简单地说，最优化的实质就是求出那些能够使目标函数达到极值的选择变量的值的集合。

举例来说，某厂商可能寻求利润 $\pi$ 最大化，即最大化总收益 $R$ 与总成本 $C$ 的差。因为在给定技术水平和市场对该厂商产品

① 有时它们也被称作决策变量或政策变量。

233需求的情况下，$R$ 与 $C$ 均为产出水平 $Q$ 的函数，因而 $\pi$ 也可以表示成 $Q$ 的函数：

$$\pi(Q)=R(Q)-C(Q)$$

此方程构成了目标函数，$\pi$ 为最大化之目标，$Q$ 则是唯一的选择变量。最优化问题是选择产出水平 $Q$ 使得 $\pi$ 最大化。注意，根据定义，$\pi$ 的最优水平即是其最大化水平，但是选择变量 $Q$ 的最优水平本身并不要求是最大值或最小值。

为将此问题纳入更一般的模式以便于进一步讨论（尽管我们这里仍局限于只有一个选择变量的目标函数），我们考察一般函数：

$$y=f(x)$$

以研究出这样一种方法：求出 $x$ 值的水平，从而使 $y$ 值最大化或最小化。在下面的讨论中，我们将假定函数 $f$ 连续可导。

## 9.2 相对极大值和极小值：一阶导数检验

因为目标函数 $y=f(x)$ 是以一般形式表示的，对其是线性的还是非线性的，是单调的还是既含有递增成分，又含有递减成分等，并无限制。在符合上述目标函数形式的诸多可能类型中，我们选择了三种类型，描绘在图 9.1 中。尽管这些图形可能很简单，但它们对我们了解确定函数 $y=f(x)$ 的极大值或极小值问题是有极大价值的。

### 相对极值与绝对极值

如图 9.1$a$ 所示，若目标函数为常函数，无论变量 $x$ 选择何

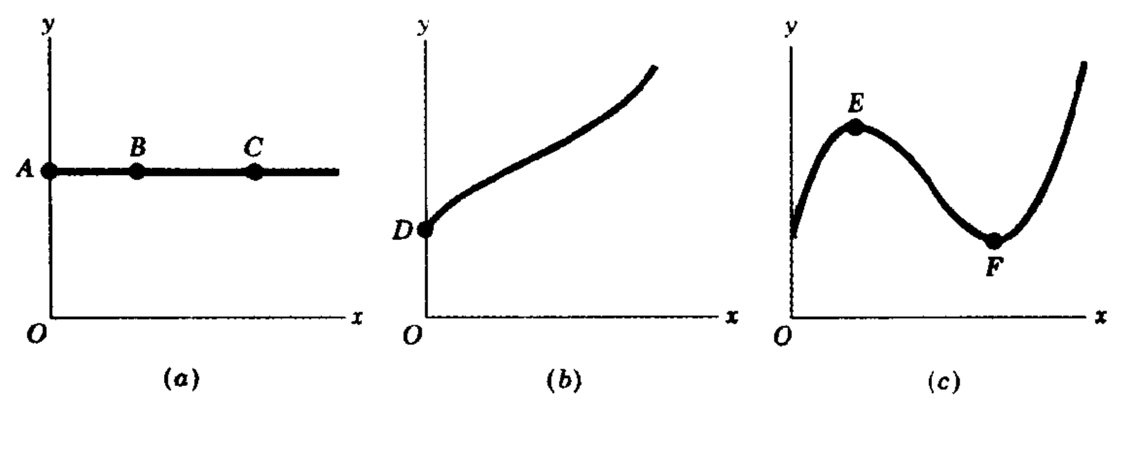



图 9.1

值，都会得到同样的 $y$ 值。该函数图形上每一点（如 $A$、$B$ 或 $C$）234 的高度都可以看作极大值或是极小值，或者既非极大值也非极小值。在此情况下，为极大化或极小化 $y$ 而对 $x$ 进行选择没有什么意义。

在图 9.1$b$ 中，函数是单调递增的，若其定义域为非负实数，那么它就没有有限的极大值。但我们可以将左端的值 $D$（$y$ 截距）看作极小值；事实上，它是函数值域中的绝对（或总体）极小值。

而图 9.1$c$ 中的点 $E$ 和 $F$ 则是相对（或局部）极值的例子，因为它们中的每个点仅代表在该点紧密邻域中的极值。当然，点 $F$ 是相对极小值并不能确保它也是函数的总体极小值，尽管在此例中它可能恰好是总体最小值。类似地，相对极大值点 $E$ 可能是，也可能不是总体极大值。还要注意的是，一个函数很可能有几个极值，其中一些是极大值，另一些是极小值。

在我们要处理的绝大多数经济问题中，我们即便不是全部，也是主要关心非终点值的极值。因为大多数此类问题的目标函数的定义域被限定为非负数的集合，因而左端终点值表示选择变量处

于零水平，所以通常没有什么实际意义。实际上在经济分析中最常遇到的函数类型是图 9.1$c$ 中的函数，或者是此类函数的变形：即曲线中仅有一处弯曲。因此，我们后面的讨论主要集中于寻找像 $E$ 点和 $F$ 点这样的相对极值。即便我们想要求出绝对极值，先求出相对极值也是必要的，因为绝对极值必定为相对极值或者终点值中的一个。因此，如果我们知道了所有相对极值，则我们只需选择其中最大的一个与终点值相比较，便可确定绝对极值。函数的绝对最小值也可以类似的方法发现。后面，我们所要考察的极值都是指相对或局部极值，除非另作说明。

## 一阶导数检验

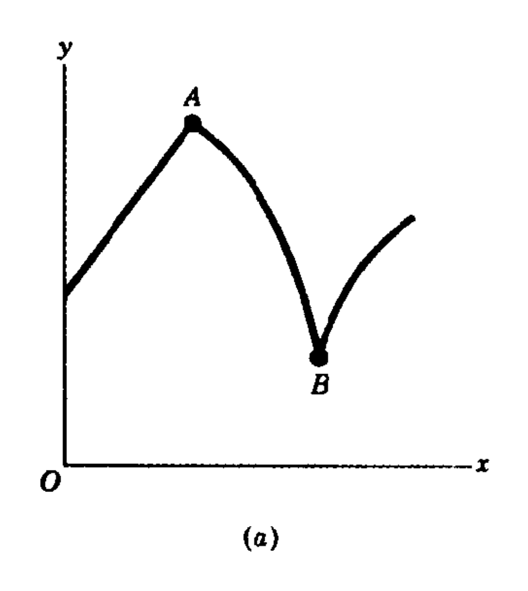
作为一个术语，今后我们把函数的导数称作一阶导数。这样称呼的理由很快就会清楚。

给定函数 $y=f(x)$，其一阶导数 $f'(x)$ 在寻求极值方面起着重要作用。这是因为若函数在 $x=x_0$ 处存在极值，则我们或者有 235 (1) $f'(x)=0$；(2) 或者 $f'(x_0)$ 不存在。第二种情况在图 9.2$a$ 中描述，其中 $A$ 点和 $B$ 点代表 $y$ 的相对极值，但在这些角点处，函数的导数不存在。然而在现在的讨论中，我们假设 $y=f(x)$ 连续，并具有连续偏导数，所以我们实际上已经排除了角点的存在。对于平滑的函数，相对极值仅在一阶导数为零处存在。图 9.2$b$ 中的 $C$ 点和 $D$ 点说明了这种情况：这两个点均表示极值，且二者均具有斜率为零的特征，即 $f'(x_1)=0$ 和 $f'(x_2)=0$。不难看出，当斜率不为零时，我们不可能得到相对极小值（谷底），或者相对极大值（峰顶）。为此，在与平滑函数相关的内容中，我们取 $f'(x)=0$ 作为相对极值（极大值或极小值）存在的必要条件。

但我们必须补充一点：斜率为零尽管是相对极值存在的必要条件，但却不是充分条件。我们很快就会给出斜率为零但却与极值无关的例子。然而，如果在导数为零的条件的基础上再增加某些附加条件，我们就会得到一种相对极值检验的重要方法。此方法可以表述如下：

### 相对极值的一阶导数检验

若函数 $f(x)$ 在 $x = x_0$ 处的一阶导数等于零，即 $f'(x) = 0$，则函数在 $x_0$ 的值 $f(x_0)$ 将是：



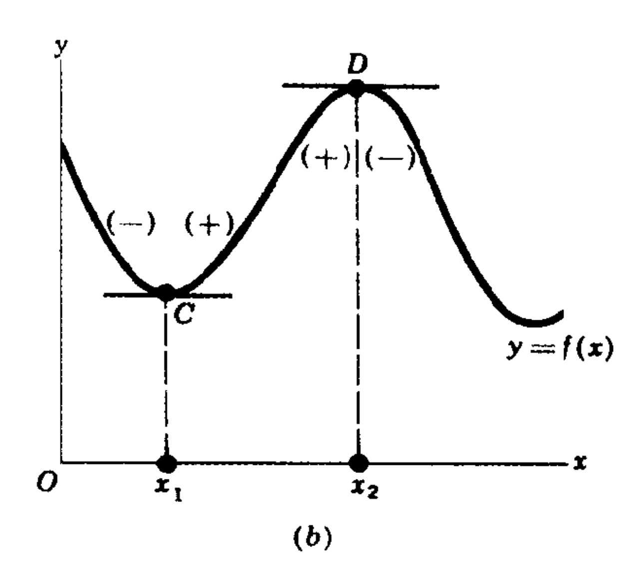




图 9.2

*a*. 当 $x$ 从 $x_0$ 左边增至 $x_0$ 右边时，若 $f'(x)$ 由正变为负，则存在相对极大值。

*b*. 若 $x$ 从 $x_0$ 的左边增至 $x_0$ 右边时，若 $f'(x)$ 由负变正，则存在相对极小值。

*c*. 当 $x$ 从 $x_0$ 左边增至 $x_0$ 右边时，若 $f'(x)$ 的符号不变，则既无相对极大值存在，也无相对极小值存在。

236 如果 $f'(x_0)=0$,则我们称 $x_0$ 为 $x$ 的临界值,称 $f(x_0)$为 $y$ 或函数 $f$ 的稳定值。相应地,把坐标为 $x_0$ 和 $f(x_0)$的点称作稳定点。(我们使用稳定一词的合理性是不言自明的:只要是在斜率为零的地方,所研究的点就不会处于向上或向下倾斜的斜面上,而总是位于相对静止的位置。)所以从几何上看,上文中所列的第一种可能的稳定点将会是在峰顶,如图 9.2*b* 中的点 $D$;而第二个可能的稳定点则会是在谷底,如图 9.2*b* 中的 $C$。但要注意,考虑到还存在第三种可能性(待讨论),我们还不能将条件 $f'(x)=0$ 视为相对极值存在的充分条件。然而,我们现在可以看到,如果必要条件 $f'(x)=0$ 得到满足,则导数符号变化的条件可以视为相对极值存在的充分条件;是极大值还是极小值取决于符号变化的方向。

现在我们来解释第三种可能性。在图 9.3*a* 中,函数 $f$ 在点 $J$(当 $x=j$ 时)的斜率为零。但是,虽然 $f'(j)=0$,$f(j)$是一个稳定值,可当 $x$ 从 $x=j$ 的一边变至其另一边时,导数的符号并不发生变化。因此,根据上面的检验方法可知,点 $J$ 既非极大值,也非极小值,从图形中我们也可以看到这一点。确切地说,点 $J$ 只不过是人们所说的拐点的一个例子而已。

拐点的特征是:在该点,导函数(与原函数相对照)达到极值。237 由于此极值可能是极大值,也可能是极小值,所以我们有两类拐点。在图 9.3*a*′中,我们绘出了导数 $f'(x)$的图形,从图中可以看到,当 $x=j$ 时(即点 $J'$),导数值等于零,但在 $J'$两边均为正,所以 $J'$是导函数 $f'(x)$的极小值。

另一类拐点绘在图 9.3*b* 中,函数 $g(x)$的斜率在 $k$ 以前递增,在 $k$ 以后递减。因而,导函数 $g'(x)$的图形绘在图 *b*′中,其中

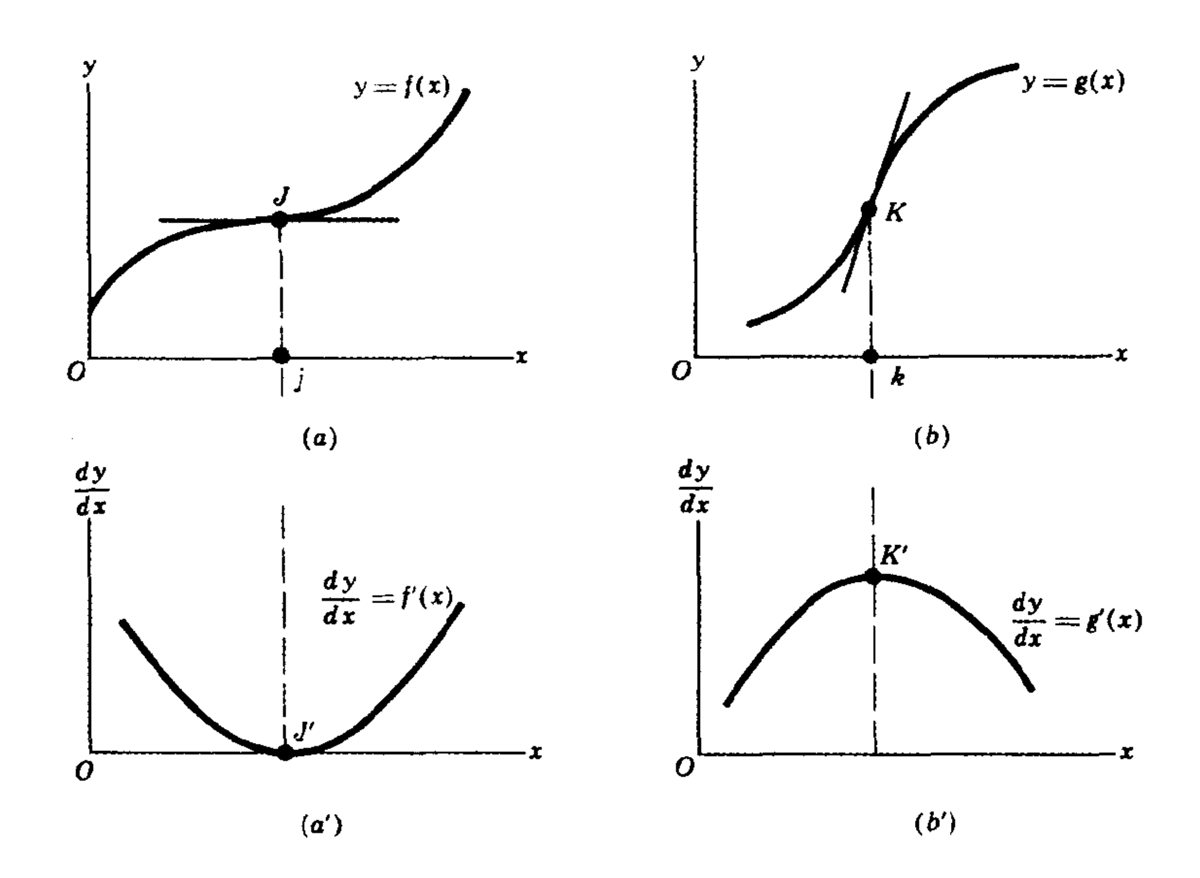



图 9.3

$K'$是导函数 $g'(x)$的极大值。①

总之,相对极值必定为稳定值,但稳定值或者与相对极值相联系,或者与拐点相联系。因此,要求出给定函数的相对极大值或极小值,第一步应先求出 $f'(x)=0$ 的函数稳定值;第二步再运用一阶导数检验法来确定每个稳定值是极大值、极小值,还是二者都不是。

**例 1** 求函数

$$y=f(x)=x^3-12x^2+36x+8$$

的相对极值。首先,求出函数的导数

$$f'(x)=3x^2-24x+36$$

① 注意,导数值为零,尽管是相对极值的必要条件,但并不是拐点的必要条件;因为导数 $g'(x)$在 $x=k$ 的值为正,但点 $K$ 是拐点。

为求出临界值,即满足条件 $f'(x)=0$ 的 $x$ 的值,我们令二次导函数等于零,得到二次方程

$$3x^2-24x+36=0$$

将其因式分解,或应用二次公式,我们得到下面两个根:

$\bar{x}_1=2$ [在该点我们有 $f'(2)=0$ 和 $f(2)=40$]

$\bar{x}_2=6$ [在该点我们有 $f'(6)=0$ 和 $f(6)=8$]

因为 $f'(2)=f'(6)=0$,这两个 $x$ 值为我们所求的临界值。

很容易验证,在 $x$ 的最小邻域内,对 $x<2$, $f'(x)>0$, $x>2$, $f'(x)<0$;因此,相应的函数值 $f(2)=40$ 是相对极大值;类似地,在 $x=6$ 的最小邻域内,因当 $x<6$ 时,$f'(x)<0$; $x>6$ 时,$f'(x)>0$,函数值 $f(6)=8$ 必然为相对极小值。

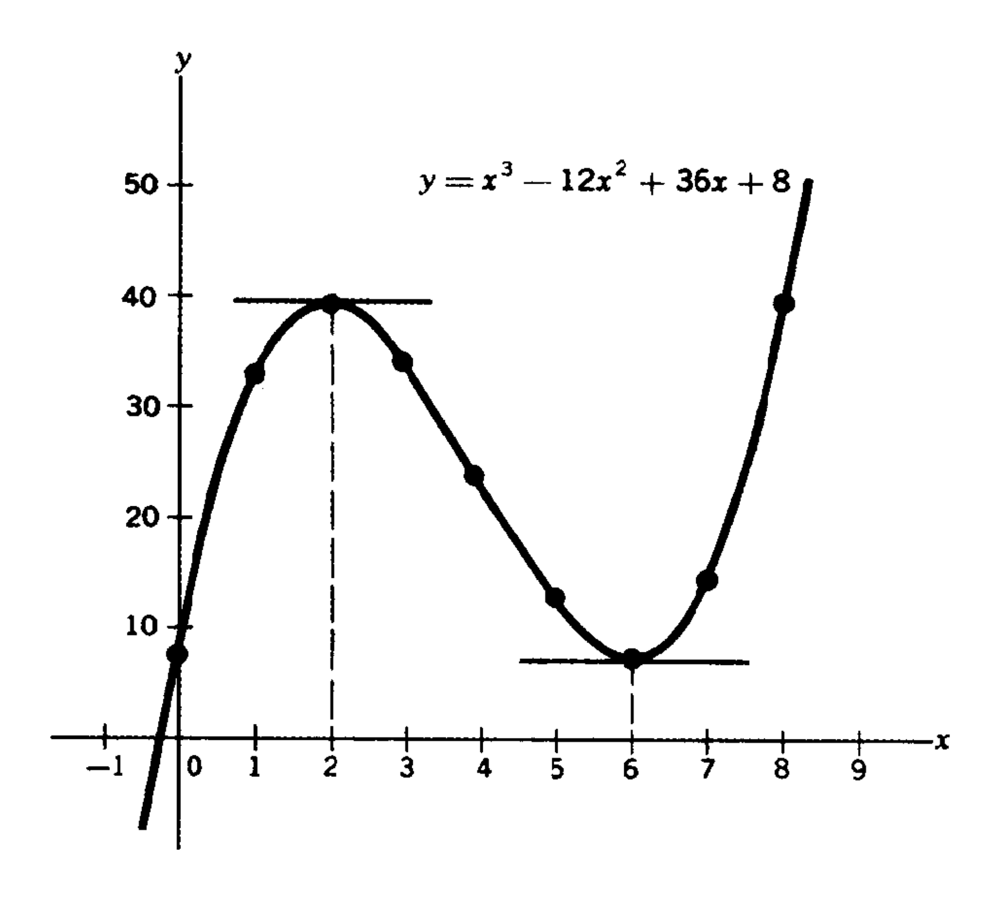



图 9.4

此例的函数图形绘在图 9.4 中。这样的图形也许有助于验证通过一阶导数检验所得到的极值的位置。但实际上，在多数情况 238 下这种“帮助”来自于相反的方向：在数学上所推导出的极值有助于绘出图形；准确地绘出图形从理论上看，需要知道函数在定义域内每个点的值；但实际上要绘出图形，只要在定义域内选择若干个点，再将其余有代表性的点插入即可。这种绘图方法的问题在于：除非由于巧合碰上了稳定点，否则我们便难以画准曲线转折点的位置。现在，掌握了一阶导数检验方法，我们便能准确地确定转折点的位置。

**例 2** 求平均成本函数的相对极值

$$AC = f(Q) = Q^2 - 5Q + 8$$

其导数 $f'(Q) = 2Q - 5$，是一个线性方程。令 $f'(Q) = 0$，得到 $2Q - 5 = 0$，此方程只有一个根 $\overline{Q} = 2.5$。在本例中只有一个临界值。要应用一阶导数检验，我们需要分别求出比如 $Q = 2.4$ 和 $Q = 2.6$ 的导数值。因 $f'(2.4) = -0.2 < 0$，而 $f'(2.6) = 0.2 > 0$，所以我们可以得出结论，稳定值 $AC = f(2.5) = 1.75$ 表示相对极小值。本例的函数图形实际上是一个 $U$ 形曲线，所以此相对极小值也是绝对极小值。确切地知道这个点的位置对描绘 $AC$ 曲线图具有极大的益处。

## 练习 9.2

1　假设定义域为全部实数的集合，求下列函数的稳定值，并检验其为 239 极大值、极小值还是拐点：

(a) $y=-2x^2+4x+9$　　(c) $y=x^2+3$

(b) $y=5x^2+x$　　(d) $y=3x^2-6x+2$

2　假设定义域为区间$[0,\infty)$，求下列函数的稳定值，并检验其为相对极小值、极大值，还是拐点：

(a) $y=x^3-3x+5$

(b) $y=\frac{1}{3}x^3-x^2+x+10$

(c) $y=-x^3+4.5x^2-6x+6$

3　证明函数 $y=x+1/x\,(x\neq 0)$ 具有两个相对极值，一个为极大值，另一个为极小值。"极小值"是大于还是小于"极大值"？为什么会出现这种相悖的结果？

4　令 $T=\phi(x)$ 为一总函数，如总生产函数、总成本函数等：

(a) 写出边际函数 $M$ 和平均函数 $A$ 的表达式。

(b) 证明：当 $A$ 达到相对极值时，$M$ 和 $A$ 的值必相同。

(c) 上述结论对在同一图形中绘出边际曲线与平均曲线有什么一般原则意义？

(d) 当总函数达到极值点 $A$ 时，你对总函数在该点的极值能得出何结论？

---

## 9.3　二阶及高阶导数

迄今为止，我们仅考察了 $y=f(x)$ 的一阶导数 $f'(x)$。现在让我们引入二阶导数以及更高阶导数的概念。这些概念有助于我们提出确定相对极值的其它标准。

### 导数的导数

因为导数 $f'(x)$ 本身是 $x$ 的函数，若它是连续和平滑的，则它对 $x$ 也应是可微的。微分的结果便是函数 $f$ 的二阶导数，表示成

$f''(x)$　其中双撇号表示函数 $f(x)$ 对 $x$ 微分两次，而双撇号后面的表达式 $(x)$ 表明二阶导数仍是 $x$ 的函数。

或者

$\frac{d^2y}{dx^2}$　这个符号源于这样一种考虑：二阶导数实际上意味着 $\frac{d}{dx}\left(\frac{dy}{dx}\right)$；因此，$d^2$ 出现在此符号的分子上，$dx^2$ 出现在分母上。

240

若 $f''(x)$ 对定义域内所有 $x$ 值都存在，则称函数 $f(x)$ 是二阶可微的。此外，若 $f''(x)$ 连续，则称 $f(x)$ 是二阶连续可微的。[①]

作为 $x$ 的函数，二阶导数可对 $x$ 再次微分得到**三阶**导数，依次还可以得到**四阶**导数；只要可微性条件得到满足，还可以得到无穷高阶导数。按照二阶导数同样的表示方法，这些导数可表示为：

$f'''(x)$，$f^{(4)}(x)$，…，$f^{(n)}(x)$ [上标加上小括号(　)]

或者

$$\frac{d^3y}{dx^3},\ \frac{d^4y}{dx^4},\ \cdots,\ \frac{d^ny}{dx^n}$$

上述最后一个表达式还可以写成 $\frac{d^n}{dx^n}y$，其中 $\frac{d^n}{dx^n}$ 部分作为算符，表示对 $x$ 取 $n$ 阶导数。

几乎我们要处理的所有的具体函数都有高至我们所期望的任

---

① 人们也常运用下面的符号来表示函数的连续性和可微性

$f\in C^{(0)}$ 或 $f\in C$：　$f$ 是连续函数

$f\in C^{(1)}$ 或 $f\in C'$：　$f$ 是连续可微的

$f\in C^{(2)}$：　$f$ 是二阶连续可微的

符号 $C^{(n)}$ 表示拥有 $n$ 阶在定义域内连续的所有函数的集合。

何阶的连续导数,即它们都连续可微任意次。无论何时使用一般函数,如 $f(x)$,我们总是假设它有可达到我们所需要的阶数的导数。

**例 1** 求下面函数的一阶至五阶导数:

$$y=f(x)=4x^4-x^3+17x^2+3x-1$$

所需求的函数如下:

$$\begin{aligned}f'(x)&=16x^3-3x^2+34x+3\\f''(x)&=48x^2-6x+34\\f'''(x)&=96x-6\\f^{(4)}(x)&=96\\f^{(5)}(x)&=0\end{aligned}$$

241 在这个特定的例子(多项式函数)中,每一个更高一阶的导数都比前一个简单,到五阶导数时,它便等于零了。但正如我们下面的例子所表明的那样,并非所有类型的导数都如此,这里应强调的是"五阶导数等于零"与"五阶导数不存在"描述的是两种完全不同的情况。还应注意的是,$f^{(5)}(x)=0$(对所有 $x$ 值都等于零)与 $f^{(5)}(x_0)=0$ (仅在 $x_0$ 为零)也是不同的。

**例 2** 求如下有理函数的一至四阶导数

$$y=g(x)=\frac{x}{1+x}\quad(x\neq-1)$$

这些导数既可以运用商求导法则求出,也可以在将函数改写成 $y=x(1+x)^{-1}$后,用积的求导法则求出:

$$\left.\begin{aligned}g'(x)&=(1+x)^{-2}\\g''(x)&=-2(1+x)^{-3}\\g'''(x)&=6(1+x)^{-4}\\g^{(4)}(x)&=-24(1+x)^{-5}\end{aligned}\right\}(x\neq-1)$$

在本例中，反复求导，并不能使更高阶的导数比前一阶导数更简单。

注意，所得到的各阶导数，像原函数 $g(x)$ 一样，也是 $x$ 的函数。给定 $x$ 的具体值，这些导数也会取具体值。例如当 $x$ 取 2 时，例 2 的二阶导数可计算为

$$g''(2) = -2(3)^{-3} = \frac{-2}{27}$$

对 $x$ 的其它值，也可以作同样的计算。最重要的是必须认识到，要像我们上面做过的那样，计算二阶导数 $g''(x)$ 在 $x=2$ 的值，我们必须首先由 $g'(x)$ 得到 $g''(x)$，再将 $x=2$ 代入方程。在对 $g'(x)$ 微分从而得到 $g''(x)$ 之前将 $x=2$ 代入函数 $g(x)$ 或 $g'(x)$ 是错误的。

## 二阶导数的解释

导函数 $f'(x)$ 度量函数 $f$ 的变化率。同样，二阶导函数 $f''$ 度量一阶导数 $f'$ 的变化率，换句话说，二阶导数度量原函数 $f$ 变化率的变化率。为说明一阶和二阶导数的区别，在 $x=x_0$ 处给出自变量 $x$ 的无穷小增量，

242

$$\left.\begin{matrix} f'(x_0)>0 \\ f'(x_0)<0 \end{matrix}\right\} \text{意味着函数值趋向于} \begin{cases} \text{递增} \\ \text{递减} \end{cases}$$

而对于二阶导数

$$\left.\begin{matrix} f''(x_0)>0 \\ f''(x_0)<0 \end{matrix}\right\} \text{意味着曲线斜率趋向于} \begin{cases} \text{递增} \\ \text{递减} \end{cases}$$

因此，在 $x=x_0$，正的一阶和二阶导数意味着曲线在该点的斜率不仅为正，而且递增，即函数值以递增的速率递增。同样，一阶导

数为正而二阶导数为负表明曲线斜率为正但递减，即函数值以递减的速率递增。负的一阶导数的情况也可以作类似的解释。但应提醒的是，当 $f'(x_0)<0$ 和 $f''(x_0)>0$ 时，曲线斜率为负且递增，但这并不意味着曲线斜率这样变化；比如从(−10)变为(−11)；相反，函数从较小的数字(−11)变为较大的数字(−10)。换句话说，当 $x$ 增加时，斜率为负的曲线的切线趋于平缓。最后，当 $f'(x_0)<0, f''(x_0)<0$ 时，曲线斜率必然为负且递减。这意味着随着 $x$ 的增加，斜率为负的曲线倾向于变得较为陡峭。

因为我们一直在讨论斜率，所以运用几何说明来继续我们的讨论是有益的。在图 9.5 所示的两条抛物线上，我们标出了 $A$、$B$、$C$、$D$、$E$、$F$ 六个点，每一个点都表明一阶、二阶导数符号的不同组合，其具体结果如下：

| 若 $x$ 取 | 导数符号为 | | 描述组合的点 |
|---|---|---|---|
| $x=x_1$ | $f'(x_1)>0$ | $f''(x_1)<0$ | 点 $A$ |
| $x=x_2$ | $f'(x_2)=0$ | $f''(x_2)<0$ | 点 $B$ |
| $x=x_3$ | $f'(x_3)<0$ | $f''(x_3)<0$ | 点 $C$ |
| $x=x_4$ | $g'(x_4)<0$ | $g''(x_4)>0$ | 点 $D$ |
| $x=x_5$ | $g'(x_5)=0$ | $g''(x_5)>0$ | 点 $E$ |
| $x=x_6$ | $g'(x_6)>0$ | $g''(x_6)>0$ | 点 $F$ |

由此我们知道，负的二阶导数(前三种情况)确实反映在倒 $U$ 形曲线或它的一部分中，因为当 $x$ 增加时，我们所讨论的曲线的斜率要求越来越小。相反，正二阶导数则反映在 $U$ 形曲线或其一部分中，因为我们所讨论的曲线的斜率必然随着 $x$ 的增加而增大。从横轴角度观察图 9.5，我们看到图 $a$ 中的那条曲线处处是凹的，而图 $b$ 中的那条曲线则处处是凸的。因为凹性和凸性用于描述曲线如何“弯曲”，所以我们现在可以预期函数的二阶导数能告诉我

们其曲线曲率的情况，正如一阶导数可以告诉我们函数的斜率一样。

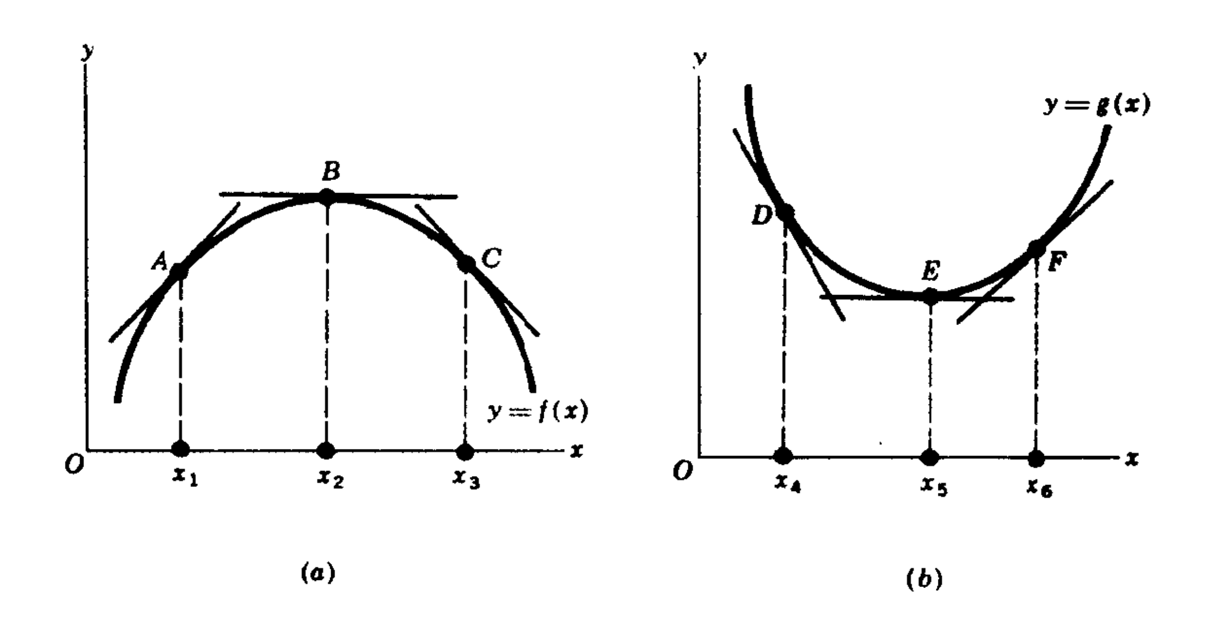



图 9.5

尽管“凹”与“凸”两个词已充分描述了图9.5中两条曲线曲率上的差别，但今天的作者都愿意更具体地将其称作严格凹和严格凸。与此相一致，图形为严格凹(严格凸)的函数称作严格凹(严格凸)函数。严格凹函数的几何特征如下：如果我们在曲线上选择任意两个点 $M$ 和 $N$，并以一条直线将它们联起来，线段 $MN$(除点 $M$ 和 $N$ 外)必定完全位于曲线的下方。如果我们将上面一段话中的“下方”改成“上方”，便得到了关于严格凸函数几何特征的描述。在图9.5中可以对此进行验证。若特征条件有所放松，使得线段 $MN$ 既可以位于曲线以下，又可以在曲线中(与曲线一致)，则我们可以去掉副词“严格”，将其称为凹函数。类似地，如果线段 $MN$ 或者位于曲线上方，或者与曲线重合，则我们还可以去掉副词“严格”，称其为凸函数。注意，因为线段 $MN$ 可与非严格凹和凸的曲线重合，因此，后者极可能包含一直线段。由此可知，严格

243

凹(凸)函数必然是凹(凸)函数,但反之不成立。[①]

根据前面对二阶导数的讨论,我们现在可以推断若对所有的 $x$,$f''(x)$为负,则原函数 $f(x)$必定为严格凹函数。类似地,若对所有的 $x$, $f''(x)$ 为正,则 $f(x)$ 必然为严格凸函数。尽管如此,将上述推断颠倒过来,比如说:若 $f(x)$ 严格凹(严格凸),则$f''(x)$必定对所有的 $x$ 都为负(正),则不成立。这是因为在某些例外情况下,二阶导数在曲线上的稳定点可以取零值。这种情况的一个例子是 $y=f(x)=x^4$,其曲线为严格凸的曲线,但其导数为:

$$f'(x)=4x^3,\ f''(x)=12x^2$$

244 它们表明,在稳定点 $x=0$,二阶导数为 $f''(0)=0$。但要注意,在任意的 $x\neq0$ 的点,此函数的二阶导数确实具有我们所期望正值。所以,除了在稳定点有可能取零值外,严格凹函数和严格凸函数一般都保持单一的代数符号。

至于其它类型的函数,二阶导数可能取正值,也可能取负值,这依赖于 $x$ 的值。例如,在图 9.3$a$ 和 $b$ 中,函数 $f(x)$和 $g(x)$的二阶导数分别在拐点 $J$ 和 $K$ 都经历了符号变化。根据图 9.3$a'$,$f'(x)$的斜率,即 $f''(x)$的值,在 $x=j$ 处由负变为正;与其恰好相反的是 $g'(x)$的斜率,即 $g''(x)$的值,其图形绘在9.3$b'$中。以描述弯曲的术语来说就是,$f(x)$的图形在点 $J$ 由凹变为凸,而 $g(x)$的图形在点 $K$ 处则有相反的变化。所以,我们可不必将拐点视为一阶导数达到极值的点,而将其赋予函数的图形在该点产生了弯曲方向上的变化,或二阶导数符号的变化的特征。

---

① 我们将在 11.5 节对这些问题进行进一步的讨论。

## 一个应用

图 9.5 给出了二次函数的一个例子，此二次函数可以如下一般形式表示：

$$y = ax^2 + bx + c \qquad (a \neq 0)$$

由我们对二阶导数的讨论，我们现在可以推出一个方便的方法来确定已知的二次函数是具有严格的凸形（$U$ 形）还是严格的凹形（倒 $U$ 形）。

因为上面的二次函数的二阶导数为 $d^2y/dx^2 = 2a$，此导数总具有与系数 $a$ 相同的代数符号。回顾一下，正的二阶导数意味着严格凸函数，由此我们可以推断在上述二次函数中系数 $a$ 若为正，将会得到 $U$ 形曲线；若系数 $a$ 为负则得到倒 $U$ 形曲线。

正如在 9.2 节末所表明的，此函数的相对极值也是其绝对极值。因为二次函数仅有一个谷底或峰顶，显然分别在 $U$ 形曲线或倒 $U$ 形曲线中。

## 练习 9.3

---

1 求下列函数的二阶和三阶导数：

(a) $ax^2 + bx + c$  (c) $\frac{2x}{1-x}$ $(x \neq 1)$

(b) $6x^4 - 3x - 4$  (d) $\frac{1+x}{1-x}$ $(x \neq 1)$

2 下列函数哪个是严格凸函数？

(a) $y = 9x^2 - 4x + 2$  (c) $u = 9 - x^2$

(b) $w = -3x^2 + 39$  (d) $v = 8 - 3x + x^2$

245　3　画出($a$)一条不是严格凹的凹曲线;($b$)一条既凹又凸的曲线。

4　已知函数 $y=a-\frac{b}{c+x}(a,\ b,\ c>0;\ x\geqslant 0)$,通过下述考察以确定函数的一般图形:

($a$)函数的一阶和二阶导数;

($b$)函数的纵截距;

($c$)当 $x$ 趋于无穷大时 $y$ 的极值。

若此函数用于表示消费函数,要使其具有经济意义,应对参数施以何种限制?

5　绘出使得 $f'(x)=0$ 的函数 $f(x)$ 的图形,以及使得 $g'(3)=0$ 的函数 $g(x)$ 的图形。用稳定点的概念以一句话来概括 $f(x)$ 与 $g(x)$ 的本质区别。

---

## 9.4　二阶导数检验

回到图 9.5 中的两个极值点 $B$ 和 $E$,并记住我们刚刚确立的二阶导数与曲线曲向之间的联系,我们应该能够看到下述判定相对极值的标准应是正确的。

**相对极值的二阶导数检验**

如果函数 $f$ 的一阶导数在 $x=x_0$ 为 $f'(x_0)=0$,且,

$a$. 若在 $x_0$ 的二阶导数值 $f''(x_0)<0$,则函数在 $x_0$ 的函数值 $f(x_0)$ 为相对极大值。

$b$. 若在 $x_0$ 的二阶导数值 $f''(x_0)>0$,则函数在 $x_0$ 的函数值 $f(x_0)$ 为相对极小值。

此检验一般来说比一阶导数检验更便于应用,因为它不要求

我们检验 $x_0$ 左右两侧的符号。但它也有一个缺陷，即在 $f''(x_0)=0$ 时无法得出确切的结论。在 $f''(x_0)=0$ 时，稳定值 $f(x_0)$ 可能是极大值，或者极小值，甚至也可能是拐点。[①] 当遇到 $f''(x_0)=0$ 的情况时，我们或者回头使用一阶导数检验，或者借助 246 于我们将在 9.6 节中提出的涉及三阶甚至更高阶导数的其它检验。

**例 1** 求如下函数的相对极值

$$y=f(x)=4x^2-x$$

其一阶和二阶导数为

$$f'(x)=8x-1 \quad 和 \quad f''(x)=8$$

令 $f'(x)=0$，解所得到的方程，我们求得(唯一的)临界值为 $\overline{x}=1/8$，由此产生(唯一的)稳定值 $f(1/8)=-1/16$。因为二阶导数为正(因为在此情况下对任意 $x$ 值，它总是为正)，此极值是最小值。实际上，因为给定函数的曲线为 $U$ 形曲线，相对极小值也是绝对最小值

**例 2** 求函数

$$y=g(x)=x^3-3x^2+2$$

的相对极值。此函数的一阶和二阶导数为

$$g'(x)=3x^2-6x \quad 和 \quad g''(x)=6x-6$$

---

① 为了解当 $f''(x_0)=0$ 时该点有可能为拐点，我们重新回到图 9.3a 和 9.3a′。上图中的点 $J$ 是拐点，$x=j$ 是其临界值。因为 $f'(x)$ 曲线在下图中的 $x=j$ 达到最小值，$f'(x)$ 的斜率即 $f''(x)$ 在临界值 $x=j$ 必定为零。因此当 $f''(x)=0$ 时点 $J$ 描述了一个拐点。

为考察相对极小值与 $f''(x_0)=0$ 也是一致的，我们考察函数 $y=x^4$。此函数可绘成 $U$ 形曲线，并在临界值 $x=0$ 达到极小值 $y=0$，因为此函数的二阶导数 $f''(x)=12x^2$，所以在临界值 $x=0$，此导数的值也为零。因此，当 $f''(x_0)=0$ 时，此函数描述了相对极小值的出现。

令 $g'(x)=0$,并解所得到的二次方程,即 $3x^2-6x=0$,我们得到临界值 $\bar{x}_1=0,\bar{x}_2=2$;由此又产生两个稳定值:

$g(0)=2$ [是极大值,因 $g''(0)=-6<0$]

$g(2)=-2$ [是极小值,因 $g''(2)=6>0$]

## 必要条件与充分条件

像在一阶导数检验中的情况一样,斜率为零的条件 $f'(x)=0$ 在二阶导数检验中充当必要条件。因为此条件以一阶导数为基础,它也常被称作一阶条件。一旦我们知道一阶条件在 $x=x_0$ 处得到满足,$f''(x_0)$ 为负(正)就成为稳定值为相对极大值(极小值)的充分条件。这些以二阶导数为基础的充分条件,通常被称二阶条件。

重复一下:一阶条件仅是相对极大值或极小值的必要条件,但不是充分条件。(还记得拐点吗?)与此形成鲜明对照的是,虽然二阶条件 $f''(x)$ 在临界值 $x_0$ 为负(正)是相对极大值(极小值)存在的充分条件,但却不是必要条件。(还记得当 $f''(x_0)=0$ 时存在相对极值吗?)因此,读者必须非常谨慎地提防进行下述推断:"因为已知稳定值 $f(x_0)$ 为极小值,所以我们必然有 $f''(x_0)>0$。"上述推理之所以错误,是因为误把 $f''(x_0)$ 为正当作 $f(x_0)$ 是极小值的必要条件。

这并不是说二阶导数永远不能作为相对极值存在的必要条件。事实上,它们是可以作为必要条件的。但必须谨慎地考虑到这一事实:相对极大值(极小值)不仅在 $f''(x_0)$ 为负(正)时存在,而且在 $f''(x_0)=0$ 时也存在。因此,二阶必要条件必须按照弱不

等式来表示：要使稳定值 $f(x_0)$ 为相对 $\begin{cases}极大值\\极小值\end{cases}$，必须使 $f''(x_0)\begin{Bmatrix}\leqslant\\\geqslant\end{Bmatrix}0$。

## 利润最大化的条件

现在我们提供一些极大值问题，即最优化问题在经济上应用的例子。

学习经济学的学生首先要学习的问题之一就是，要实现利润最大化，厂商必须使边际成本等于边际收益。我们给出这一条件的数学推导。为使我们的分析保持一般性，我们将使用总收益函数 $R=R(Q)$ 和总成本函数 $C=C(Q)$，二者均是单一变量 $Q$ 的函数。由此可知，利润函数(目标函数)也可以表示成 $Q$(选择变量)的函数：

$$\pi = \pi(Q) = R(Q) - C(Q) \tag{9.1}$$

为求得利润最大化的产出水平，必须满足最大化的一阶必要条件：$d\pi/dQ=0$。因此，将(9.1)对 $Q$ 微分，并令所得的导数等于零。结果为：

$$\begin{aligned}\frac{d\pi}{dQ} &\equiv \pi'(Q) = R'(Q) - C'(Q)\\ &= 0 \quad 当且仅当\ R'(Q) = C'(Q)\end{aligned} \tag{9.2}$$

因此，**最优**产出(**均衡**产出) $\overline{Q}$ 必须满足方程 $R'(\overline{Q})=C'(\overline{Q})$，或者 MR = MC。此条件构成了利润最大化的一阶必要条件。

然而，满足一阶条件也可能得到最小值，而不是最大值，因此下一步我们还得检验二阶条件。将(9.2)对 $Q$ 求导得到二阶导数：

$$\frac{d^2\pi}{dQ^2} \equiv \pi''(Q) = R''(Q) - C''(Q)$$

$$<0 \text{ 当且仅当 } R''(Q) < C''(Q)$$

248 对于使得 $R'(\bar{Q}) = C'(\bar{Q})$的产出水平 $\bar{Q}$,满足二阶条件 $R''(\bar{Q}) < C''(\bar{Q})$是使其成为利润最大化产出水平的充分条件。从经济的角度看,这意味着在 MC = MR 的产出水平时,如果 MR 的变化率低于 MC 的变化率,则该产出将使利润最大化。

这些条件描述在图 9.6 中。在图 *a* 中,我们绘出了总收益和总成本曲线。这两条曲线在产出水平 $Q_2$ 和 $Q_4$ 相交了两次。在开区间$(Q_2, Q_4)$中,总收益 $R$ 超过总成本 $C$,因此 $\pi$ 为正。但在区间$[0, Q_2)$ 和 $(Q_4, Q_5]$,其中 $Q_5$ 代表厂商生产能力的上限,$\pi$ 为负。这一事实反映在图 *b* 中,其中的利润曲线——它根据每一产出水平 $R$ 曲线与 $C$ 曲线之间的垂直距离绘出——仅在区间$(Q_2, Q_4)$位于横轴的上方。

与一阶条件相一致,当我们令 $d\pi/dQ = 0$ 时,我们的目的便是要确定利润曲线在产出 $Q_3$ 的峰点 $K$,此点的斜率为零。但是在产出 $Q_1$ 的相对极小值点 $M$ 也是一个候选点,因为它也满足斜率为零的要求。我们后面要借助于二阶条件来排除"错误"的极值。

一阶条件 $d\pi/dQ = 0$ 与条件 $R'(Q) = C'(Q)$是等价的。在图 9.6*a* 中,产出水平 $Q_3$ 满足这一条件,因为 $R$ 与 $C$ 曲线在 $Q_3$ 斜率相同(在 $H$ 点与 $J$ 点绘出的两条切线是平行的),在产出水平 $Q_1$ 也同样如此。因为 $R$ 与 $C$ 曲线斜率相同意味着 MR = MC,产出水平 $Q_3$ 和 $Q_1$ 必定为 MR 曲线与 MC 曲线的交点,如图 9.6*c* 所示。

二阶条件是如何发挥作用的呢?我们首先看图 9.6*b*。在点 $K$,$\pi$ 函数的二阶导数的值为负(除掉零值的例外情况),即$\pi''(Q_3)$

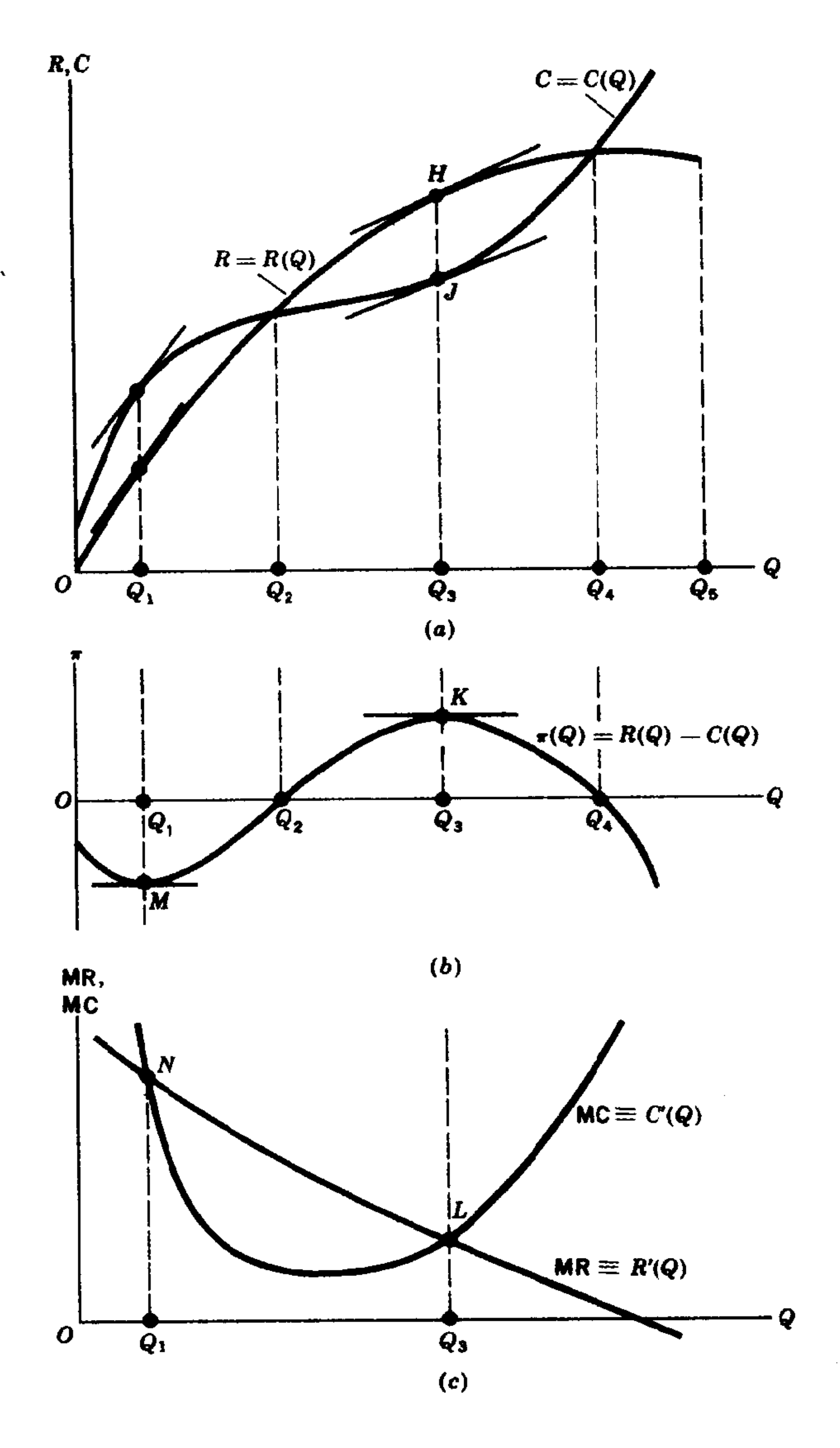

R, C
C = C(Q)
H
R = R(Q)
J
O
Q
$Q_1$
$Q_2$
$Q_3$
$Q_4$
$Q_5$
(a)
π
K
π(Q) = R(Q) − C(Q)
O
Q
$Q_1$
$Q_2$
$Q_3$
$Q_4$
M
(b)
MR,
MC
N
MC ≡ C′(Q)
L
MR ≡ R′(Q)
O
Q
$Q_1$
$Q_3$
(c)

图 9.6

$<0$，因为曲线在 $K$ 点附近为倒 $U$ 形，这意味着 $Q_3$ 将使利润最大化。另一方面，在点 $M$，我们预期 $\pi''(Q_1)>0$，$Q_1$ 使 $\pi$ 达到极小值，当然极大值的二阶充分条件也可以表述为 $R''(Q)<C''(Q)$，即 MR 曲线的斜率小于 MC 曲线的斜率。由图 9.3$c$ 立即可以看到产出 $Q_3$ 可以满足这一条件，因为在点 $L$，MR 的斜率为负，而 MC 的斜率为正。但产出 $Q_1$ 不符合这一条件，因为此处 MC 与 MR 的斜率均为负，而且在点 $N$，MR 斜率的数值小于 MC 的斜率，这意味着 $R''(Q)$ 大于 $C''(Q)$。所以实际上，产出 $Q_1$ 也不符合相对极大值的二阶必要条件，但满足相对极小值的二阶充分条件。

**例 3** 令 $R(Q)$ 与 $C(Q)$ 函数为

$$R(Q)=1200Q-2Q^2$$

$$C(Q)=Q^3-61.25Q^2+1528.5Q+2000$$

250 则利润函数为

$$\pi(Q)=-Q^3+59.25Q^2-328.5Q-2000$$

其中 $R$、$C$、$\pi$ 均以美元为单位，而 $Q$ 则可以吨/周为单位。此函数有两个临界值 $Q=3$ 和 $Q=36.5$，因为

$$\frac{d\pi}{dQ}=-3Q^2+118.5Q-328.5=0 \quad \text{当 } Q=\begin{cases}3\\36.5\end{cases}$$

但因二阶导数为

$$\frac{d^2\pi}{dQ^2}=-6Q+118.5 \quad \begin{cases}>0 & \text{当 } Q=3\\<0 & \text{当 } Q=36.5\end{cases}$$

所以利润最大化产出为 $\overline{Q}=36.5$（吨/周）。（另一产出使利润最小。）通过把 $\overline{Q}$ 代入利润函数，我们可以求得最大利润为

$$\overline{\pi}=\pi(36.5)=16318.44 \text{（美元/周）}$$

本例还可用另一办法求解。我们可以先求出 MR 与 MC 函

数，然后令二者相等，即求其交点。因为

$$R'(Q) = 1200 - 4Q$$

$$C'(Q) = 3Q^2 - 122.5Q + 1528.5$$

令两个函数相等，将得到一个与 $d\pi/dQ = 0$ 相一致的二次方程，它能产生两个与前面一样的临界值。

## 三次总成本函数的系数

在上面例 3，我们使用了三次函数来表示总成本函数。如图 9.6*a* 所示，传统的总成本曲线 $C = C(Q)$ 被假定含有两个“扭动”，从而形成一个凹孤（递减的边际成本）和一个凸弧（递增的边际成本）。因为三次函数，如图 9.4 所示，总是含有两次波动，所以它能够很好地充当这一角色。但图 9.4 还提醒我们警惕一个问题：三次函数的图形可能产生一个向下倾斜的弧段，而要使总成本函数具有经济意义，它必须始终保持向上倾斜（更大的产出要承担更高的成本）。如果我们要运用的成本函数为

$$C = C(Q) = aQ^3 + bQ^2 + cQ + d \qquad (9.3)$$

则有必要对参数施以适当的限制，以防止 $C$ 曲线向下倾斜。

描述上述要求的另一种方式是：MC 函数必须处处为正；仅当 MC 函数的**绝对极小值**为正时，才可以保证这一点。将(9.3)对 $Q$ 微分，我们得到 MC 函数

$$\text{MC} = C'(Q) = 3aQ^2 + 2bQ + c \qquad (9.4)$$

因为它是一个二次函数，所以其图形如图 9.6*c* 中的抛物线。要使 MC 曲线处处为正（位于横轴之上），必须使抛物线为 $U$ 形（换言之，为倒 $U$ 形。）因此，(9.4)中的 $Q^2$ 项的系数必须为正，即我们必须限定 $a>0$。然而，此限制绝非充分条件，因为 $U$ 形 MC 曲线

的最小值——称其为 $MC_{min}$(一个相对极小值,也可能恰好是绝对极小值)——仍可能出现于横轴之下。因此我们下一步必须求出 $MC_{min}$,并确定使其为正所需的参数限制。

根据我们关于相对极值的知识,MC 的相对极小值将出现于

$$\frac{d}{dQ}MC = 6aQ + 2b = 0$$

满足一阶条件的产出水平是

$$Q^* = \frac{-2b}{6a} = \frac{-b}{3a}$$

它使 MC 最小化(而非最大化),因为考虑到 $a>0$,二阶导数 $d^2(MC)/dQ^2 = 6a$ 一定为正、知道了 $Q^*$,现在我们就可以计算 $MC_{min}$了,但我们现在也可以由它来推断系数 $b$ 的符号。因为可排除负的产出水平,所以我们知道 $b$ 不能为正(给定 $a>0$)。进而,因为在某一正的产出水平,边际收益递减规律成立,即 MC 在初始阶段下降,$Q^*$应为正而不是为零。因此,我们必然有约束 $b<0$。

现在将 MC 最小的产出 $Q^*$代入(9.4)以求得 $MC_{min}$就很简单了

$$MC_{min} = 3a\left(\frac{-b}{3a}\right)^2 + 2b\,\frac{-b}{3a} + c = \frac{3ac-b^2}{3a}$$

因此,为保证 $MC_{min}$为正,我们还必须施加约束[①] $b^2<3ac$。这最

---

① 此约束也可以通过完全平方法得到。MC 函数可连续变换如下:

$$\begin{aligned} MC &= 3aQ^2 + 2bQ + c \\ &= \left(3aQ^2 + 2bQ + \frac{b^2}{3a}\right) - \frac{b^2}{3a} + c \\ &= \left(\sqrt{3a}Q + \sqrt{\frac{b^2}{3a}}\right)^2 + \frac{-b^2+3ac}{3a} \end{aligned}$$

因为平方表达式有可能为零,在知道 $a>0$ 的情况下,仅当 $b^2<3ac$ 时,MC 为正才能得到保证。

后一个约束,实际上也意味着约束 $C>0$(为什么?)。

上述讨论涉及三个参数 $a$、$b$、$c$。那么另一个参数 $d$ 呢? 答 252 案是对 $d$ 也需要加以限制,但它与保持 MC 为正无关。若我们令(9.3)中的 $Q=0$,我们求得 $C(0)=d$。因此,$d$ 的作用仅在于决定 MC 曲线的纵截距,而与其斜率无关。因为 $d$ 的经济意义是固定成本,所以适当的约束(仅在短期)应当是 $d>0$。

总之,总成本函数(9.3)的系数应被施加如下限制(仅与短期有关):

$$a,\ c,\ d>0 \quad b<0 \quad b^2<3ac \tag{9.5}$$

正如你能够验证的那样,例 3 中的 $C(Q)$函数确实满足(9.5)。

## 向上倾斜的边际收益曲线

在图 9.6$c$ 中边际收益曲线被表示成一条处处向下倾斜的曲线,当然,这也是在不完全竞争条件下厂商 MR 曲线的传统画法。但是,绝不能事先排除 MR 曲线部分,甚至全部向上倾斜的可能性。[①]

给定平均收益函数 $\text{AR}=f(Q)$,边际收益函数可以表示成

$$\text{MR}=f(Q)+Qf'(Q) \qquad [\text{由}(7.7)]$$

MR 曲线的斜率则可由下述导数确定 253

$$\frac{d}{dQ}\text{MR}=f'(Q)+f'(Q)+Qf''(Q)=2f'(Q)+Qf''(Q)$$

只要 AR 曲线向下倾斜(正如在不完全竞争市场中那样),则

① 这一点在约翰·P.丰比,斯蒂芬·拉森和詹姆斯·史密斯的论文“需求定律、正斜率的边际收益以及多重利润均衡”中被重点揭示出来。此论文载于《经济探索》,1982 年 4 月,第 303—311 页。

$2f'(Q)$必定为负。但$Qf''(Q)$可能为负，或者是零，也可能为正，这取决于AR函数二阶导数的符号，即取决于AR曲线是严格凹的、线性的，还是严格凸的。如果AR曲线是严格凸的，无论是全部（如图7.2所描述的那样），还是沿着一个特定的弧段，那么，（正的）$Qf''(Q)$大于（负的）$2f'(Q)$，从而导致MR曲线全部或部分向上倾斜的可能性是存在的。

**例4** 令平均收益函数为

正如可验证的那样（见练习9.4—7），此函数给出向下倾斜的AR曲线，所以适于不完全竞争下的厂商。因为

$$AR=f(Q)=8000-23Q+1.1Q^2-0.018Q^3$$

由此得MR的斜率为

$$MR=f(Q)+Qf'(Q)=8000-46Q+3.3Q^2-0.072Q^3$$

因为这是一个二次方程，且因$Q^2$的系数为负，所以如图9.5所示，$dMR/dQ$必定可以对$Q$绘成一个倒$U$形曲线。若此曲线的一部分恰好位于横轴之上，则MR的斜率将取正值。

令$MR/dQ=0$并应用二次公式，我们求得二次函数的两个根为$Q_1=10.76$，$Q_2=19.79$（近似值）。这意味着对于在开区间（$Q_1$，$Q_2$）的$Q$值，$dMR/dQ$曲线确实位于横轴之上。所以在$Q_1$和$Q_2$之间的产出水平上，边际收益曲线的斜率确实为正。

MR曲线上存在正斜率的弧段具有有趣的含义。外形较多的弯曲MR曲线可能会与MC曲线产生不止一个满足利润最大化二阶充分条件的交点。然而，尽管这些交点构成了局部最优，但仅有一个是厂商追求的全局最优。

## 练习 9.4

1 以二阶导数检验,求 $y$ 的相对极大值和极小值:

($a$) $y=-2x^2+8x+25$

($b$) $y=x^3+6x^2+7$

($c$) $y=\frac{1}{3}x^3-3x^2+5x+3$

($d$) $y=\frac{2x}{1-2x}$ $(x\neq\frac{1}{2})$

2 格林萨姆先生想在他的房子边设计一个长方形花坛,另外三边以金属网围上。他只有 32 英尺金属网可用。当矩形的长 $L$ 和宽 $W$ 为多少时,才能使种植面积最大?你如何保证你给出的答案能得到最大面积,而不是最小面积?

3 某厂商有如下总成本与总需求函数:

$$C=\frac{1}{3}Q^3-7Q^2+111Q+50$$
$$Q=100-P$$

($a$) 此总成本函数是否满足(9.5)的系数限制?

($b$) 写出以 $Q$ 表示的总收益函数 $R$。

($c$) 列出以 $Q$ 表示的总利润函数 $\pi$。

($d$) 求出利润最大化的产出水平 $\bar{Q}$。

($e$) 最大利润是多少?

4 若(9.3) 中的系数 $b$ 为零,边际成本曲线与总成本曲线将会如何?

5 二次利润函数 $\pi(Q)=hQ^2+jQ+k$ 用于下列假设: 254

($a$) 若什么也不生产,由于固定成本的关系,利润将为负。

($b$) 利润函数为严格凹函数。

($c$) 在正产出水平 $\overline{Q}$ 达到利润最大化。这要求对参数进行何种限制?

6 某完全竞争厂商有单一可变投入 $L$(劳动),每期工资率为 $W_0$。若该厂商每期的固定成本为 $F$ 美元,产品的价格为 $P_0$:

($a$) 写出厂商的生产函数、收益函数、成本函数和利润函数。

($b$) 何为利润最大化的一阶条件?解释此条件的经济意义。

($c$) 什么样的经济环境才能保证利润最大而不是最小?

7 用下列步骤验证例 4 中的 AR 曲线具有负的斜率:

($a$) 以 $S$ 标示 AR 的斜率;写出一个 $S$ 的表达式。

($b$) 运用二阶导数检验,求 $S$ 的极大值,即 $S_{max}$。

($c$) 则由 $S_{max}$的值可推断出 AR 曲线的斜率为负。

---

## 9.5 关于麦克劳林级数与泰勒级数的题外讨论

现在我们到了开发一种新的相验方法的时候了。即便二阶导数在稳定点的值为零时,这种方法也可以应用。但在此之前,我们有必要先讨论如何将函数 $y=f(x)$分别"展开"为我们所知道的麦克劳林级数(在点 $x=0$ 附近展开)和泰勒级数(在任意点 $x=x_0$ 附近展开)的问题。

在目前的内容中,在 $x_0$ 附近展开函数 $y=f(x)$意味着把此函数变换成一个多项式,其中各项系数均以导数 $f'(x_0)$,$f''(x_0)$等来表示——所有导数均在展开点 $x_0$ 处计算其值。在麦克劳林级数中,均在 $x=0$ 处计算各阶导数的值,所以在系数中我们有 $f'(0)$、$f''(0)$等。展开的结果可以称为幂级数,因它是由幂函数构成的多项式。

### 多项式函数的麦克劳林级数

我们首先考察 $n$ 次多项式函数

$$f(x)=a_0+a_1x+a_2x^2+a_3x^3+a_4x^4+\cdots+a_nx^n \tag{9.6}$$

的展开。因为这涉及将一个多项式变换为另一个多项式,似乎这

是一种毫无意义、毫无目的的练习，但它实际上能够清楚地阐明函数展开的思想。

因为展开后的幂级数包含函数 $f$ 的各阶导数，所以我们首先 255 求出这些导数。对(9.6)连续求导，我们得到以下各阶导数：

$$
\begin{aligned}
f'(x) &= a_1 + 2a_2x + 3a_3x^2 + 4a_4x^3 + \cdots + na_nx^{n-1} \\
f''(x) &= 2a_2 + 3(2)a_3x + 4(3)a_4x^2 + \cdots \\
&\quad + n(n-1)a_nx^{n-2} \\
f'''(x) &= 3(2)a_3 + 4(3)(2)a_4x + \cdots \\
&\quad + n(n-1)(n-2)a_nx^{n-3} \\
f^{(4)}(x) &= 4(3)(2)a_4 + 5(4)(3)(2)a_5x + \cdots \\
&\quad + n(n-1)(n-2)(n-3)a_nx^{n-4} \\
&\vdots \\
f^{(n)}(x) &= n(n-1)(n-2)(n-3)\cdots(3)(2)(1)a_n
\end{aligned}
$$

注意，每次逐次求导，项数就减少一个——加在前面的常数项消失了——直至 $n$ 阶导数，剩下一个常数项(乘积项)。这些导数值可在不同的 $x$ 值处计算；这里我们将在 $x=0$ 处计算，结果所有含有 $x$ 的项将消失。于是我们得到如下非常整洁的导数值：

$$
\begin{aligned}
&f'(0) = a_1 \quad f''(0) = 2a_2 \quad f'''(0) = 3(2)a_3 \\
&f^{(4)}(0) = 4(3)(2)a_4 \cdots \\
&f^{(n)}(0) = n(n-1)(n-2)(n-3)\cdots \\
&\qquad (3)(2)(1)a_n
\end{aligned}
\tag{9.7}
$$

若我们现在采用简写符号 $n!$ (读作：$n$ 的阶乘)，定义为

$$n! = n(n-1)(n-2)\cdots(3)(2)(1) \quad (n = \text{正整数})$$

因而如 2! =2×1=2, 3! =3×2×1=6 等。(0! 定义为等于1),则结果(9.7)可重写成:

$$a_1=\frac{f'(0)}{1!} \quad a_2=\frac{f''(0)}{2!} \quad a_3=\frac{f'''(0)}{3!} \quad a_4\frac{f^{(4)}(0)}{4!}$$

$$\cdots\ a_n=\frac{f^{(n)}(0)}{n!}$$

将其代入(9.6)并利用 $f(0)=a_0$ 这一明显的事实,我们现在可以将已知函数 $f(x)$ 表示成一个新的多项式,其中的系数用 $x=0$ 时的导数值表示:①

$$f(x)=\frac{f(0)}{0!}+\frac{f'(0)}{1!}x+\frac{f''(0)}{2!}x^2+$$

$$\frac{f'''(0)}{3!}x^3+\cdots+\frac{f^{(n)}(0)}{n!}x^n \tag{9.8}$$

这个新的多项式,即多项式函数 $f(x)$ 的麦克劳林级数,表示函数 $f(x)$ 围绕零($x=0$)展开。

256 **例 1** 求

$$f(x)=2+4x+3x^2 \tag{9.9}$$

的麦克劳林级数。此函数有导数

$$f'(x)=4+6x \qquad \text{则}\begin{cases} f'(0)=4 \\ f''(0)=6 \end{cases}$$
$$f''(x)=6$$

麦克劳林级数为

$$f(x)=f(0)+f'(0)x+\frac{f''(0)}{2}x^2$$

---

① 因为 0! =1 和 1! =1,所以(9.8)式等号右侧的前两项可以简单地分别写成 $f(0)$ 和 $f'(0)x$,我们这里在分母中写上 0! 和 1! 是为了引起读者对展开式中不同项间的对称性的注意。

$$= 2 + 4x + 3x^2$$

这验证了麦克劳林级数确实正确地表示了给定的函数。

## 多项式函数的泰勒级数

更一般地，多项式函数(9.6)可在任意点 $x = x_0$ 附近展开，而不一定非在零处展开。为简化起见，我们利用具体的二次函数(9.9)来对此加以解释，然后再将此结果一般化。

为了围绕一个特定点 $x_0$ 展开函数，我们可先把任意给定的 $x$ 值看成异于 $x_0$ 的点。更具体地，可令 $x = x_0 + \delta$，其中 $\delta$ 表示与 $x_0$ 的偏差。根据这些解释，已知函数(9.9)及其导数现在则成为：

$$\begin{aligned} f(x) &= 2 + 4(x_0 + \delta) + 3(x_0 + \delta)^2 \\ f'(x) &= 4 + 6(x_0 + \delta) \\ f''(x) &= 6 \end{aligned} \qquad (9.10)$$

我们知道，表达式$(x + \delta) = x$ 在函数中是一个变量，但在目前情况下，$x_0$ 是一个固定值，在(9.10)中仅有 $\delta$ 可视为变量。因此，$f(x)$实际上是 $\delta$ 的函数，比如可表示为 $g(\delta)$：

$$g(\delta) = 2 + 4(x_0 + \delta) + 3(x_0 + \delta)^2 \quad [\equiv f(x)]$$

其导数为

$$\begin{aligned} g'(\delta) &= 4 + 6(x_0 + \delta) \qquad && [\equiv f'(x)] \\ g''(\delta) &= 6 && [\equiv f''(x)] \end{aligned}$$

我们已知道如何在零$(\delta = 0)$的某邻域内展开 $g(\delta)$。根据(9.8)，如此展开将得到麦克劳林级数：

$$g(\delta) = \frac{g(0)}{0!} + \frac{g'(0)}{1!}\delta + \frac{g''(0)}{2!}\delta^2 \qquad (9.11)$$

但因我们令 $x = x_0 + \delta$，所以 $\delta = 0$ 意味着 $x = x_0$，因此在恒等式 257

$g(\delta)\equiv f(x)$的基础上，对 $\delta=0$ 的情况，我们可以写成

$$g(0)=f(x_0) \quad g'(0)=f'(x_0) \quad g''(0)=f''(x_0)$$

将结果代入(9.11)，我们求得表示函数 $f(x)$在 $x_0$ 附近展开的结果。因为其系数包含导数 $f'(x_0)$，$f''(x_0)$等，这些导数均在 $x=x_0$处计值：

$$f(x)[=g(\delta)]=\frac{f(x_0)}{0!}+\frac{f'(x_0)}{1!}(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2 \qquad (9.12)$$

读者可将此结果，即 $f(x)$的泰勒级数与(9.11)中 $g(\delta)$的麦克劳林级数进行比较。

因为对于所考察的具体函数(9.9)，我们有：

$$f(x_0)=2+4x_0+3x_0^2 \quad f'(x_0)=4+6x_0 \quad f''(x_0)=6$$

泰勒公式(9.12)将产生

$$f(x)=2+4x_0+3x_0^2+(4+6x_0)(x-x_0)+\frac{6}{2}(x-x_0)^2$$

$$=2+4x+3x^2$$

这验证了泰勒级数确实正确表示了给定函数。

展开式(9.12)可推广应用于(9.6)的 $n$ 次多项式。一般化的泰勒级数公式为

$$f(x)=\frac{f(x_0)}{0!}+\frac{f'(x_0)}{1!}(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n \qquad (9.13)$$

此式与(9.8)的麦克劳林级数的差别仅在于 $x_0$ 取代零作为展开点，以及以$(x-x_0)$取代了 $x_0$(9.13)告诉我们，已知 $n$ 次多项式

$f(x)$,如果我们在(9.13)右边的项中(比如)令 $x=7$,选择任意数 $x_0$,则计算这些项的值并将其相加,我们将准确地得到 $f(7)$,即 $f(x)$在 $x=7$ 的值。

**例 2** 取 $x_0=3$ 作为展开点,我们可将(9.6)等价地写成

$$f(x) = f(3) + f'(3)(x-3) + \frac{f''(3)}{2}(x-3)^2 + \cdots + \frac{f^{(n)}(3)}{n!}(x-3)^n$$

## 任意函数的展开

迄今为止,我们已说明了如何将一个 $n$ 次多项式以另一种形式的 $n$ 次多项式来表示。正如要证明的那样,只要任意函数 $\phi(x)$(不必一定是多项式)在展开点 $x_0$ 具有有限、连续和直至所需阶数的导数,就有可能将其以类似于(9.13)的多项式形式表示。258

根据被称之为泰勒定理的数学命题,给定任意函数 $\phi(x)$,若我们知道此函数在 $x=x_0$ 的值 $\phi(x_0)$和其在 $x_0$ 的导数值$\phi'(x_0)$、$\phi''(x_0)$等等,则此函数可在点 $x_0$ 的邻域展开如下($n$ 为任意选定的正整数):

$$\phi(x) = \left[\frac{\phi(x_0)}{0!} + \frac{\phi'(x_0)}{1!}(x-x_0) + \frac{\phi''(x_0)}{2!}(x-x_0)^2 + \cdots + \frac{\phi^{(n)}(x_0)}{n!}(x-x_0)^n\right] + R_n \qquad (9.14)$$

$$\equiv P_n + R_n$$

其中 $P_n$ 表示括号内的前$(n+1)$项 $n$ 次多项式,$R_n$表示余项;[1]

[1] 不要把符号 $R_n$(remainder) 与符号 $R^n$($n$ 维空间相混淆)。

余项这个概念下面再解释。$R_n$ 的存在使得(9.14)有别于(9.13),因此,(9.14)被称作带有余项的泰勒级数。多项式 $P_n$ 的类型与余项 $R_n$ 的大小取决于我们如何选择 $n$ 值。$n$ 值越大,$P_n$ 中的项数越多;相应地,对每一个不同的 $n$,$R_n$ 也有不同的值。这个事实说明了给这两个符号加上下标 $n$ 的必要性。为便于记忆,我们可以把 $n$ 视为 $P_n$ 中最高阶导数的阶数。(在 $n=0$ 的特殊情况下,$P_n$ 中根本不会出现导数。)

在(9.14)中出现 $R_n$,是由于我们这里讨论的是任意函数 $\phi$,它不能总是被精确地变换为(9.13)所示的多项式形式。因此,它包括一个余项作为 $P_n$ 的补充,以表示 $\phi(x)$与多项式 $P_n$ 的区别。从另一个角度看,$P_n$ 可视为一个近似于 $\phi(x)$的多项式,而 $R_n$ 则是近似误差的度量。例如,若我们选择 $n=1$,则我们有

$$\phi(x)=[\phi(x_0)+\phi'(x_0)(x-x_0)]+R_1=P_1+R_1$$

其中 $P_1$ 包括 $n+1=2$ 项,并构成了 $\phi(x)$的线性近似。若 $n=2$,则会出现 2 次幂的项,所以

$$\begin{aligned}\phi(x)&=\left[\phi(x_0)+\phi'(x_0)(x-x_0)+\frac{\phi''(x_0)}{2!}(x-x_0)^2\right]+R_2\\&=P_2+R_2\end{aligned}$$

其中 $P_2$ 包括 $n+1=3$ 项,它是 $\phi(x)$的二次近似,等等。

259 我们应当顺便提一句,任意函数 $\phi(x)$显然应当包括 $n$ 次多项式(9.6)。当然这是它的一个特例。在后一种情况下,如果将其展开成一个 $n$ 次多项式,则(9.13)的结果完全可以应用。换句话说,我们可以应用(9.14)的结果,但 $R_n=0$,但若给定的 $n$ 次多项式展开为一个低次多项式,则后者只能视为 $f(x)$的近似,且余项会出现;相应地,(9.14)中的结果只能应用非零的余项。因此,

(9.14)类型的泰勒级数是非常一般化的。

**例3** 在点 $x_0=1$ 的某邻域内展开如下非多项式函数，取 $n=4$：

$$\phi(x)=\frac{1}{1+x}$$

我们需求出 $\phi(x)$ 的前四阶导数

$$\phi'(x)=-(1+x)^{-2} \qquad 从而\phi'(1)=-(2)^{-2}=\frac{-1}{4}$$

$$\phi''(x)=2(1+x)^{-3} \qquad \phi''(1)=2(2)^{-3}=\frac{1}{4}$$

$$\phi'''(x)=-6(1+x)^{-4} \qquad \phi'''(1)=-6(2)^{-4}=\frac{-3}{8}$$

$$\phi^{(4)}(x)=24(1+x)^{-5} \qquad \phi^{(4)}(1)=24(2)^{-5}=\frac{3}{4}$$

我们还可以看到 $\phi(1)=\frac{1}{2}$。因此，令(9.14)中的 $x_0=1$，并运用上面推导的结果，我们得到如下带有余项的泰勒级数：

$$\begin{aligned}\phi(x)&=\frac{1}{2}-\frac{1}{4}(x-1)+\frac{1}{8}(x-1)^2-\frac{1}{16}(x-1)^3\\&\quad+\frac{1}{32}(x-1)^4+R_4\\&=\frac{31}{32}-\frac{13}{16}x+\frac{1}{2}x^2-\frac{3}{16}x^3+\frac{1}{32}x^4+R_4\end{aligned}$$

当然，选择 $x_0=0$ 作为展开点也是可以的。在这种情况下，令(9.14)中的 $x_0=0$，展开的结果将是带有余项的麦克劳林级数。

**例4** 在 $x_0=1$ 附近展开下列二次函数，取 $n=1$：

$$\phi(x)=5+2x+x^2$$

像例1中(9.9)式一样，此函数为一个二次多项式，但我们的任务却是将其展开成一次多项式($n=1$)，即求出给定二次函数的线性 260 近似，所以，必然会出现余项。因此，为进行泰勒展开，将 $\phi(x)$ 视

为一个“任意”函数。

为完成这一展开，我们仅需一阶导数 $\phi'(x)=2+2x$。在 $x_0=1$ 处计值，给定函数及其导数产生

$$\phi(x_0)=\phi(1)=8 \qquad \phi'(x_0)=\phi'(1)=4$$

因此，带有余项的泰勒级数为

$$\begin{aligned}\phi(x) &= \phi(x_0)+\phi'(x_0)(x-x_0)+R_1\\ &= 8+4(x-1)+R_1=4+4x+R_1\end{aligned}$$

其中 $(4+4x)$ 项表示线性近似，$R_1$ 项则代表近似误差。

在图 9.7 中，$\phi(x)$ 的图形是一个抛物线，而其线性近似，则是 $\phi(x)$ 曲线在点(1,8)的切线。切点在 $x=1$ 处出现并非由于巧合；相反，它是展开点确定在那个特定的值 $(x_0=1)$ 的直接结果。这

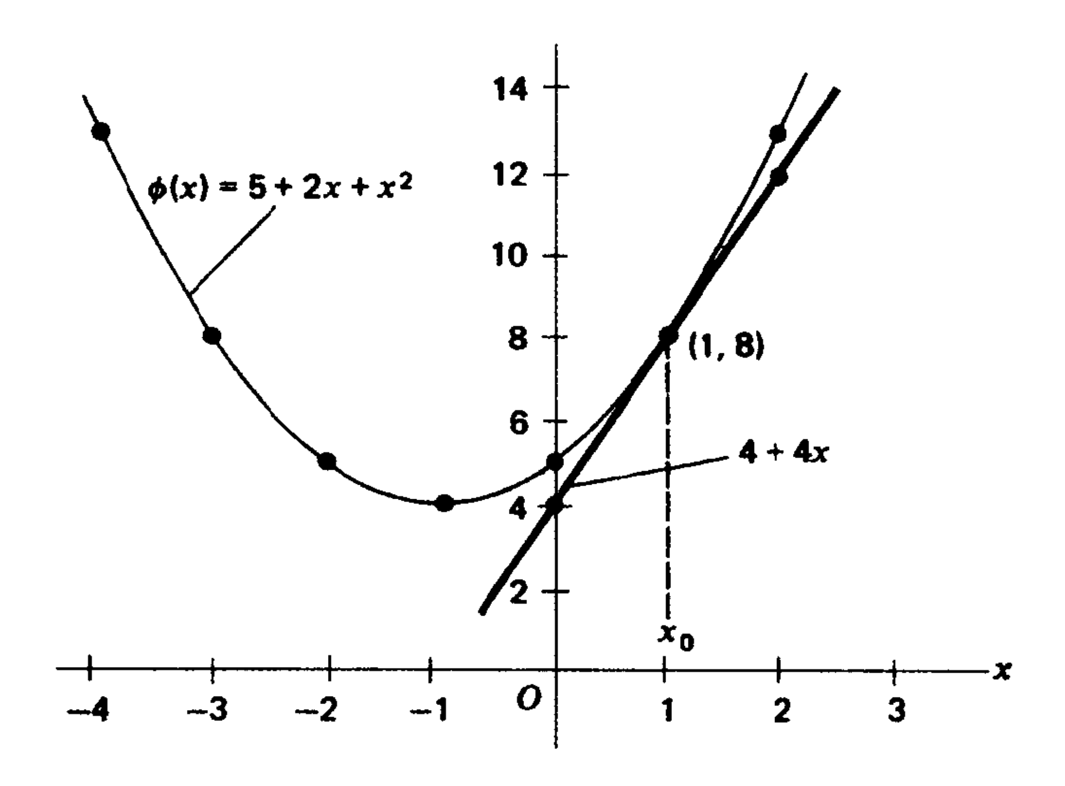



图 9.7

表明，当任意函数 $\phi(x)$ 被近似地表示成一个多项式时，后者将给

出 $\phi(x)$ 在展开点（但仅仅是在展开点）的确切值，其误差为零，($R_1=0$)。在其它点，$R_1$ 严格不等于零，而且当近似 $\phi(x)$ 的 $x$ 值越来越远离展开点 $x_0$ 时，$R_1$ 则反映了不断增大的近似误差。

## 余项的拉格朗日型

261

现在我们必须对余项作进一步讨论。根据余项的拉格朗日型，我们可以将 $R_n$ 表示为：

$$R_n = \frac{\phi^{(n+1)}(p)}{(n+1)!}(x-x_0)^{n+1} \tag{9.15}$$

其中 $P$ 为是位于 $x$（我们要计算任意函数 $\phi$ 的值的点）与 $x_0$（我们展开函数 $\phi$ 的点）之间的某个数。注意，这个表达式与(9.14)中从逻辑上看紧接着 $P_n$ 中最后一项的那一项非常相似，只是这里涉及的导数在点 $P$ 而不是在 $x_0$ 计值。此外因为点 $P$ 没有设定，所以我们不能依据此式计算 $R_n$；不过它确实具有很重要的分析意义。所以，我们以图形来解释其含义，尽管我们仅能对 $n=0$ 的简单情况加以说明。

当 $n=0$ 时，在多项式 $P_0$ 中不会出现任何导数，因此，(9.14)简化为

$$\phi(x) = P_0 + R_0 = \phi(x_0) + \phi'(p)(x-x_0)$$

或 $\quad \phi(x) - \phi(x_0) = \phi'(p)(x-x_0)$

这个结果——亦即中值定理的简化形式——表明函数 $\phi$ 在 $x_0$ 的值与在任意其它 $x$ 处的值之差可以表示为差 $(x-x_0)$ 与在 $p$（$p$ 是 $x$ 与 $x_0$ 间的某点）计值的导数 $\phi'$ 的积。让我们观察图 9.8，其中函数 $\phi(x)$ 是一条在各点均有导数的连续曲线。令 $x_0$ 为选定的展开点，$x$ 为横轴上的任意点，若我们以 $\phi(x_0)$ 近似表示 $\phi(x)$，或

262

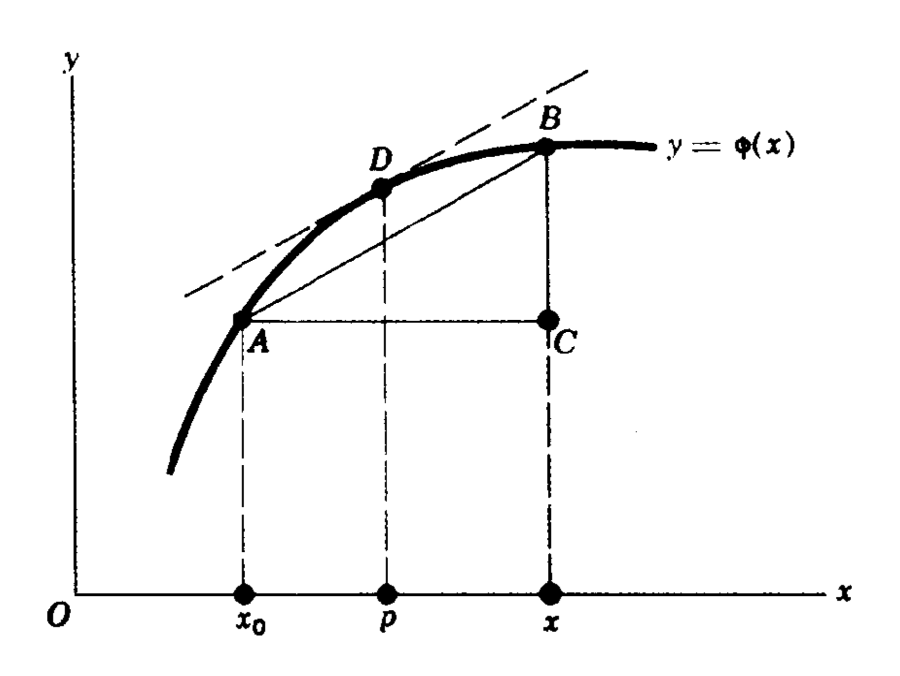



图 9.8

以距离 $x_0A$ 近似表示 $xB$，其误差将为 $\phi(x)-\phi(x_0)$，或者距离 $CB$。中值定理的含义是：误差 $CB$（即展开式中余项 $R_0$ 的值）可以表示成 $\phi'(p)(x-x_0)$，其中 $p$ 为 $x$ 与 $x_0$ 间的某点。首先，在曲线上 $A$ 点与 $B$ 点之间，我们确定点 $D$，使得过点 $D$ 的切线平行与 $AB$；因为由 $A$ 到 $B$ 的曲线是连续、平滑的，所以这个 $D$ 点必定存在。则余项将为

$$
\begin{aligned}
R_0 = CB &= \frac{CB}{AC}\cdot AC = (AB \text{ 的斜率})\cdot AC \\
&= (D \text{ 点切线的斜率})\cdot AC \\
&= (\text{曲线在 } x = p \text{ 的斜率})\cdot AC \\
&= \phi'(p)(x-x_0)
\end{aligned}
$$

其中 $p$ 点正如所要求的那样，位于 $x$ 与 $x_0$ 之间。这证明了在 $n=0$ 情况下余项的拉格朗日型的合理性。我们总可以将 $R_0$ 表示成 $\phi'(p)(x-x_0)$，因为尽管不能确定 $p$ 的具体值，但我们可以肯定，这样的点是存在的。

方程(9.15)提供了一种表示余项 $R_n$ 的方式，但它不能排除 $R_n$ 作为 $\phi(x)$ 与多项式 $P_n$ 之间偏差的根源。然而，若恰好有

当 $n \to \infty$ 时，$R_n \to 0$ 使得

当 $n \to \infty$ 时，$P_n \to \phi(x)$

则通过选择足够大的 $n$ 值，即在多项式 $P_n$ 中包括足够多的项，[①] 就可以使 $P_n$ 对 $\phi(x)$ 的近似达到我们需要的精度。在这种(方便的)情况下，我们说泰勒级数在该展开点**收敛于** $\phi(x)$。我们将于后面10.2节中讨论一个这样的例子。

## 练习 9.5

---

1 求下列阶乘的值：

(a) 5!　(b) 7!　(c) $\frac{4!}{3!}$　(d) $\frac{6!}{3!}$　(e) $\frac{(n+2)!}{n!}$

2 求下列函数麦克劳林级数的前五项(取 $n=4$，并令 $x_0=0$)

(a) $\phi(x)=\frac{1}{1-x}$　(b) $\phi(x)=\frac{1-x}{1+x}$

3 求上题两个函数的泰勒级数，取 $n=4$，$x_0=-2$。

4 根据带有拉格朗日型余项的泰勒级数[见(9.14)和(9.15)]，证明在展开点($x=x_0$)的泰勒级数总是给出在该点的**精确**值 $\phi(x_0)$，而不仅仅是近似值。

---

① 这使我们想起在5.7节中讨论的，通过近似求逆矩阵的方法。

263

## 9.6 一元函数相对极值的 $n$ 阶导数检验

在当 $n \to \infty, R_n \to 0$ 的情况下，将函数展开成泰勒（或麦克劳林）级数，是一种很有用的近似计算方法，但我们现在关心的是应用这种展开式导出相对极值的一般检验方法。

### 泰勒展开式与相对极值

作为导出这种检验方法的预备性步骤，我们将相对极值重新定义如下：

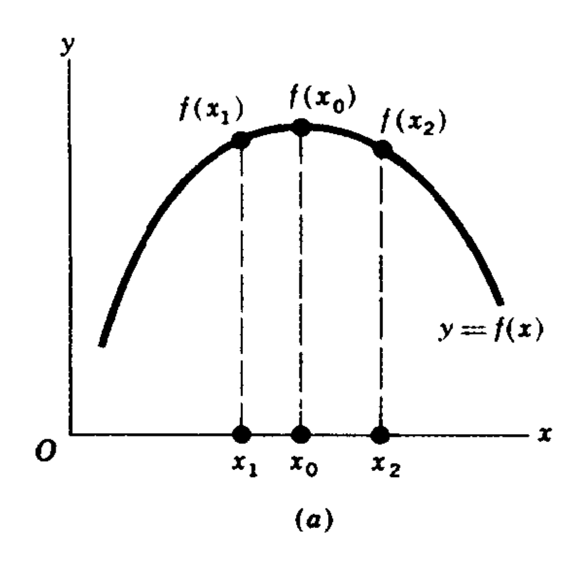
对于在 $x_0$ 最近邻域内的 $x$ 值（包括 $x_0$ 左右两边的 $x$ 值），如果 $f(x)-f(x_0)$ 为负（正），则函数 $f(x)$ 达到极大（极小）值。

参考图 9.9 会使这一定义更清楚。在图中，$x_1$ 是 $x_0$ 左边的 $x$ 值，$x_2$ 是 $x_0$ 右边的 $x$ 值。在图 $a$ 中，$f(x_0)$ 是相对极大值，因此 $f(x_0)$ 大于 $f(x_1)$ 和 $f(x_2)$。简言之，对在 $x_0$ 的最近邻域内的任意 $x$ 值，$f(x)-f(x_0)$ 为负。图 $b$ 中给出的相反的情形也同样成立，其中 $f(x_0)$ 是相对极小值，因此，$f(x)-f(x_0)>0$。

假设 $f(x)$ 在点 $x=x_0$ 具有有限的，至我们所需阶数的连续导数，则函数 $f(x)$（不必一定是多项式）可在 $x_0$ 的邻域内展开为泰勒级数。在(9.14)的基础上（仅需将 $\phi$ 变为 $f$），并运用余项的拉格朗日型，我们可写出：

$$f(x)-f(x_0)=f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\cdots$$

$$+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+\frac{f^{(n+1)}(p)}{(n+1)!}(x-x_0)^{n+1} \quad (9.16)$$



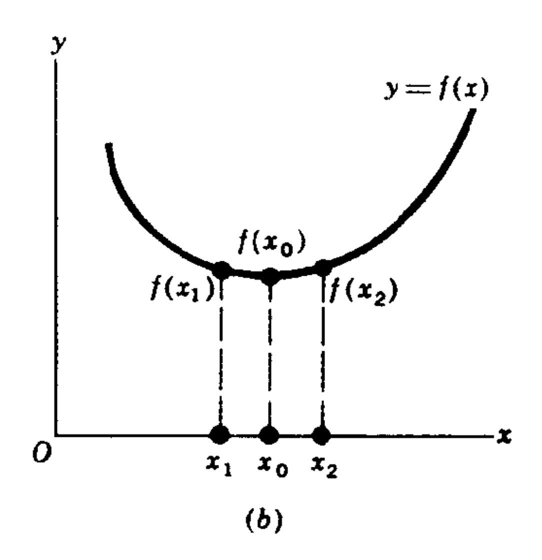




图 9.9

如果对 $x_0$ 最近邻域内的 $x$ 值,我们可以确定表达式 $f(x)-f(x_0)$的符号,那么对 $f(x_0)$是否是相对极值,若是,是相对极大值还是极小值,我们便可得出结论。为此,有必要考察(9.16)右边的和。在和中共有 $n+1$ 项,其中 $P_n$ 中有 $n$ 项,再加上一个余项,因此,实际的项数是不确定的,它取决于我们选定的 $n$ 值。但通过选择适当的 $n$,我们总可以保证右边只有一项;这样,计算 $f(x)-f(x_0)$的符号,确定 $f(x_0)$是否是极值;若是,属于哪类极值等便极大地简化了。 264

## 某些特例

通过一些具体说明,以 $n$ 阶导数检验相对极值的方法就会变得更清楚。

**特例 1**　$f'(x_0)\neq 0$

若在 $x_0$ 处一阶导数不为零,我们选择 $n=0$,则在右边仅有 $n+1=1$ 项,即只有余项 $R_0$。即我们有

$$f(x)-f(x_0)=\frac{f'(p)}{1!}(x-x_0)=f'(p)(x-x_0)$$

其中 $p$ 是在 $x_0$ 与 $x_0$ 的最近邻域中的 $x$ 值之间的一个数。注意，$p$ 与 $x_0$ 必定非常非常地接近。

右边表达式的符号是什么？因为导数的连续性，$f'(p)$将与 $f'(x_0)$具有相同的符号，因为如前所述，$p$ 必定非常非常接近于 $x_0$。在目前情况下，$f'(p)$必定为非零。事实上，它必定为某一具体的正数或负数。当我们从 $x_0$ 的左边移至 $x_0$ 的右边，$x$ 从 $x_1 < x_0$ 的数值变为 $x_2 > x_0$ 的数值(见图 9.8)，$(x-x_0)$部分会如何呢？显然，当移动时，$(x-x_0)$必然由负变为正，而且 $f(x)-f(x_0)=f'(p)(x-x_0)$也会从 $x_0$ 左边的符号变成 $x_0$ 右边的符号。但这与相对极值的新定义不符，所以当 $f'(x_0)\neq 0$ 时，在 $f(x_0)$处不存在相对极值。实际上这是我们早已知道的事实。

**特例 2** $f'(x_0)=0$；  $f''(x_0)\neq 0$

在此情况下，选择 $n=1$，从而右边起初有 $n+1=2$ 项。但因 $f'(x_0)=0$，所以其中一项消失了。这样我们仍仅有一项要计算：

$$f(x)-f(x_0)=f'(x_0)(x-x_0)+\frac{f''(p)}{2!}(x-x_0)^2$$

$$=\frac{1}{2}f''(p)(x-x_0)^2 \quad [\text{因为 } f'(x_0)=0]$$

265 同前面一样，$f''(p)$与 $f''(x_0)$有相同的符号，这个符号是确定的不变的；而$(x-x_0)^2$ 是一个平方项，总是为正。因此，表达式 $f(x)-f(x_0)$必与 $f''(x_0)$符号相同。根据上面的相对极值定义，将确定

若 $f''(x_0)<0 \quad [f'(x_0)=0]$,则 $f(x_0)$为 $f(x)$的相对极大值;

若 $f''(x_0)>0 \quad [f'(x_0)=0]$,则 $f(x_0)$为 $f(x)$的相对极小值。

读者会认出它是前面介绍的二阶导数检验。

**特例 3** $f'(x_0)=f''(x_0)=0$,但 $f'''(x_0)\neq 0$

这里我们遇到了二阶导数检验不能处理的情况,因为 $f''(x_0)$现在为零。但借助于泰勒级数,仍不难得到结论。

我们选择 $n=2$,则右边起初有三项。但因 $f'(x_0)=f''(x_0)=0$,所以其中的两项会消失,所以我们仍只需计算一项:

$$\begin{aligned} f(x)-f(x_0) &= f'(x_0)(x-x_0)+\frac{1}{2}f''(x_0)(x-x_0)^2 \\ &\quad +\frac{1}{3!}f'''(p)(x-x_0)^3 \\ &= \frac{1}{6}f'''(p)(x-x_0)^3 \end{aligned}$$

同以前一样,因导数的连续性及 $p$ 非常接近于 $x_0$,$f'''(p)$的符号与 $f'''(x_0)$的符号是一致的。但$(x-x_0)^3$ 的符号则会变化。具体地说,当 $x$ 在 $x_0$ 左边,$x-x_0$ 为负时,$(x-x_0)^3$ 也将为负,$x$ 在 $x_0$ 右边,$(x-x_0)^3$ 将为正。同样,当我们通过 $x_0$ 点时,$f(x)-f(x_0)$的符号会变化,与相对极值的定义不符。但我们知道,$x$ 是一个临界值$[f'(x_0)=0]$,由于它不是相对极值,所以必然是一个拐点。

**特例 4** $f'(x_0)=f''(x_0)=\cdots=f^{(N-1)}(x_0)=0$,但 $f^{(N)}(x_0)\neq 0$

这是一种非常一般的情况,所以我们可由此得到一个一般化的结果。注意,除 $n$ 阶导数外,其余导数均为零。

类似于上述三个特例,特例 4 的泰勒级数可简化为

$$f(x)-f(x_0)=\frac{1}{N!}f^{(N)}(p)(x-x_0)^N$$

同样，$f^{(N)}(p)$与$f^{(N)}(x_0)$的符号相同且不变。而$(x-x_0)^N$部分的符号，若$N$是奇数(参见特例1和3)，它就变化；若$N$是偶数(参见特例2)，它就保持不变。相应地，当$N$为奇数时，当我们通过点$x_0$时，$f(x)-f(x_0)$的符号就会改变，因此与相对极值的定义相违(这表明$x_0$将给出一个拐点)。但当$N$为偶数，$x$自$x_0$的左边变至右边时，$f(x)-f(x_0)$符号不变，这就确定了稳定值$f(x_0)$为相对极值，是极大值还是极小值则取决于$f^{(N)}(x_0)$为负还是正。

266

## N 阶导数检验

最后，我们可给出如下一般检验：

**一元函数相对极值的 N 阶导数检验** 如果$f(x)$在$x_0$的一阶导数$f'(x_0)=0$，如果在各阶导数中所遇到的第一个非零导数值是$N$阶导数，即$f^{(N)}(x_0)\neq 0$，则：

*a*. 若$N$为偶数，且$f^{(N)}(x_0)<0$，则稳定值$f(x_0)$将为相对极大值；

*b*. 若$N$为偶数，但$f^{(N)}(x_0)>0$，则稳定值$f(x_0)$将为相对极小值；

*c*. 若$N$为奇数，则稳定值$f(x_0)$为拐点。

上述阐述表明，当且仅当函数$f(x)$在临界值$x_0$或早或晚会产生非零导数时，才能应用$N$阶导数检验。尽管存在不满足上述条件的例外情况，但我们可能遇到的大多数函数在其各阶导数中

确实都能产生非零的 $f^{(N)}(x_0)$。[①] 因此，在绝大多数情况下，此检验都是可应用的。

**例 1** 考察函数 $y=(7-x)^4$ 的相对极值。我们取 $x=7$ 作为检验的临界值，$y=0$ 作为函数的稳定值，因为 $f'(x)=-4(7-x)^3$ 为零，当 $x=7$ 时。通过逐次求导，直至在点 $x=7$ 遇到非零导数值，我们得到：

267

$$f''(x)=12(7-x)^2 \qquad \text{从而} \qquad f''(7)=0$$
$$f'''(x)=-24(7-x) \qquad f'''(7)=0$$
$$f^{(4)}(x)=24 \qquad f^{(4)}(7)=24$$

因为 4 是偶数，且因 $f^{(4)}(7)$ 为正，所以我们确定点 $(7,0)$ 是相对极小值。

很易验证，此函数的图形为严格凸函数。由于在 $x=7$，二阶导数为零，而非为正，此例可用于说明我们前面关于二阶导数和曲线曲向的论述。其大意是：尽管对所有 $x$，正的 $f''(x)$ 意味着严格凸的 $f(x)$，但严格凸的 $f(x)$ 函数并不意味着对所有 $x$，$f''(x)$ 为

---

① 例如，若 $f(x)$ 是一个常函数，则显然 $f'(x)=f''(x)=\cdots=0$，所以永远不会得到非零导数值。但这是一种无关紧要的情况，因为常函数无需进行任何极值检验。作为一种重要情况的例子，考察函数

$$y=\begin{cases} e^{-1/x^2} & (\text{因为 } x\neq 0) \\ 0 & (\text{因为 } x=0) \end{cases}$$

其中 $y=e^{-1/x^2}$ 是一个有待于在第 10 章中介绍的指数函数。由函数本身可知，$y=e^{-1/x^2}$ 在 $x=0$ 处不连续，因为 $x=0$ 不在定义域内（被零除无定义）。但因 $\lim_{x\to 0} y=0$，通过附加规定对 $x=0$，$y=0$，则可填补定义域中的这一缺口，因而得到连续函数。其图形表明，在 $x=0$ 处函数达到极小值。但已证明在 $x=0$，高至任意阶的导数的值均为零。因此我们不能应用 $N$ 阶导数检验来确认在图形上确定的函数在 $x=0$ 处有极值这一事实。关于指数函数情况的进一步讨论，请参阅 R. 考伦特《微积分》（由 E. J. 麦克沙恩翻译）。《交叉科学》，1937 年第 1 卷 2 月号，纽约，第 196、197 及 336 页。

正。更重要的是,它还可以用于描述这一事实;给定一条严格凸(严格凹)曲线,在此曲线上求得的相对极值必定为相对极小(极大)值,因为这样的极值或者满足二阶充分条件,或者不能满足另一个的极小值(极大值)的(更高阶)的充分条件。

## 练习 9.6

---

1 求下列函数的稳定值:

$(a)\ y = x^3 \quad (b)\ y = -x^4 \quad (c)\ y = x^6 + 5$

通过 $N$ 阶导数检验确定他们是表示相对极大值、相对极小值,还是拐点。

2 求下列函数的稳定值:

$(a)\ y = (x-1)^3 + 16 \quad (b)\ y = (x-2)^4 \quad (c)\ y = (3-x)^6 + 7$

运用 $N$ 阶导数检验确定稳定值的确切性质。

---

# 第 10 章　指数函数与对数函数

上一章我们提出的 $N$ 阶导数检验，使我们得以确定任意目标 268 函数的相对极值——只要此函数仅包含一个选择变量，拥有至我们所需阶数的导数，且在临界值 $x_0$ 迟早会产生一个非零导数。但在第 9 章中，我们所举的例子仅是多项式函数和有理函数。对于这些函数，我们知道如何求出所需的导数。假设我们的目标函数恰好是一个指数函数，比如

$$y = 8^{x - \sqrt{x}}$$

则我们便无法应用导数检验标准，因为我们还需学习如何对这样函数求导。这正是本章的任务。

指数函数及与其有密切联系的对数函数，在经济学中，特别是在增长问题、一般经济动态问题中有非常重要的应用。但在本书现在这部分内容相关的具体应用则涉及一类优化问题。在这类优化问题中，时间是选择变量。例如某葡萄酒商拥有一批酒的存货。由于酿造年代的关系，葡萄酒的价值以某种特定的方式随时间的增长而增长。考虑到投入在酒的存货中的货币资本的利息成本之后，酒商的问题就是根据酒的价值函数确定销售存货的最佳时间。指数函数可以以两种方式被导入这类问题。首先，酒的价值可能随时间以某指数增长规律增长。在此情况下，我们便会有一个酒 269 的价值的指数函数。当然，这仅仅是一种可能，而非必然。然而当我们考虑到利息成本时，由于复利因素的关系（下面要解释），就规

定了指数函数导入优化问题的明确方式。因此,在我们考察优化问题之前,必须首先研究指数函数的性质。

因为我们的基本目的是处理时间作为选择变量的问题,所以现在我们改用符号 $t$(代替 $x$)表示后面讨论中的自变量。(但是,符号 $t$ 也可以很好地表示时间以外的变量。)

## 10.1 指数函数的性质

正如在与多项式相关的内容中介绍的那样,指数概念表示变量自乘的幂指标。在如 $x^3$ 或 $x^5$ 这样的幂表达式中,指数为常数;但我们没有理由不能使用变量指数,比如 $3^x$、$3^t$ 等,其中数值3被自乘至可变幂数($x$ 的不同值)。自变量作为指数而出现的函数,称作指数函数。

### 简单指数函数

简单的指数函数可通过如下形式表示:

$$y = f(t) = b^t \qquad (b > 1) \tag{10.1}$$

其中 $y$ 和 $t$ 分别表示因变量与自变量,而 $b$ 则表示指数函数不变的底,此函数的定义域为全体实数的集合。因此,与多项式函数中的指数不同,(10.1)中的可变指数 $t$ 并不限定为正整数——除非我们希望施加这种限制。

但为什么要限定 $b>1$ 呢?解释如下:因函数(10.1)的定义域为全体实数之集合,$t$ 可能取这样的值,比如 1/2。若允许 $b$ 为负,则 $b$ 的 1/2 次幂便涉及到取负数的平方根。虽然这并非不可能,但我们当然愿意采取简单的办法,限定 $b>0$。可一旦我们限定 $b>0$,

这便与限定 $b>1$ 是一致的:约束 $b>1$ 与 $b>0$ 的区别在于前者排除了(1) $0<b<1$ 和(2) $b=1$ 这两种情况;但正如我们要证明的那样,第一种情况可与限制 $b>1$ 归为一类,而第二种情况则可完全不予考虑。首先考察第一种情况:若 $b=1/5$,则我们有

$$y=\left(\frac{1}{5}\right)^t=\frac{1}{5^t}=5^{-t}$$

270

这表明有分数底的函数可以很容易重写成具有大于 1 的底的函数。至于第二种情况,事实上 $b=1$ 将给我们函数 $y=1^t=1$,所以指数函数实际上退化成为常数,不再具备指数家族成员的资格。

## 图形形式

指数函数(10.1)的图形的一般形状绘在图 10.1 中。此图是以 $b=2$ 作出的,但即便对其它 $b$ 值,图形的一般结构是相同的。

这类指数曲线的几个显著特征应加以注意。首先,它是处处连续、平滑的,因而函数是处处可微的。事实上,它可连续可微任意次。其次,它是单调递增的,而且 $y$ 始终以一个递增的速率递增,所以函数 $y=b^t$ 的一阶导数和二阶导数为正(导出相关的微分公式后,我们再对这一事实加以认证)。第三,应注意,尽管函数的定义域既包含正数,亦包含负数,但函数的值域仅限于开区间$(0,\infty)$,即无论自变量 $t$ 的符号如何,因变量 $y$ 始终为正。

指数函数的单调性至少包含两点有趣且重要的含义:首先,我们可以推断,指数函数必有一个反函数,且反函数也是单调的。我 271 们将要求出的这个反函数就是对数函数。其次,因为单调性意味着对每个给定的 $y$ 值有一个唯一的 $t$ 值与之对应,且因指数函数的值域为开区间$(0,\infty)$,由此可知我们可将任意正数表示成底

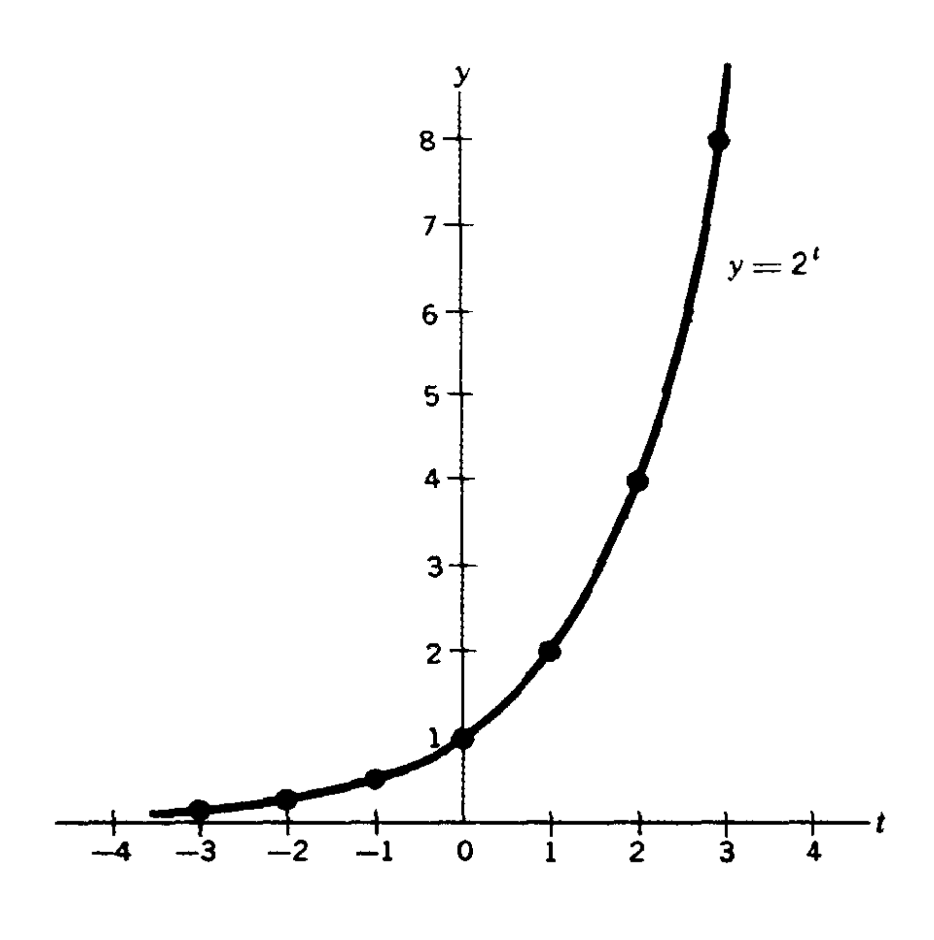



图 10.1

$b>1$的唯一的幂。这一点可由图 10.1 得到证明，图中 $y=2^t$ 的曲线包括了 $y$ 的值域中所有的正值，因此任意正的 $y$ 值必定可以表示成为 2 的某个唯一的幂。实际上，即使将底变成大于 1 的其它任意实数，只要值域不变，则仍可能将任意正数 $y$ 表示成任意底 $b>1$的一个幂。

## 推广了的指数函数

上面最后一点值得特别注意。若正值 $y$ 确实可表示成不同底的幂，则必定存在一种换底的一般方法。例如在函数 $y=9^t$ 的情况下，我们完全可以将其变换成为 $y=(3^2)^t=3^{2t}$，假设指数可以由 $t$ 变为 $2t$，则便可以将底由 9 变为 3。由换底所导致的指数变化，不会

产生任何新的函数类型，因为若我们令 $w=2t$，则 $y=3^{2t}=3^{w}$，仍与(10.1)形式一致。但是从底 3 的角度看，指数现在是 $2t$ 而非 $t$。在指数 $t$ 前增加了数字系数(这里是 2)会产生何种影响呢？

在图 10.2$a$ 中可找到答案，图中绘出两条曲线，一条是函数 $y=f(t)=b^{t}$，另一条是函数 $y=g(t)=b^{2t}$。因后一函数的指数恰为前者之 2 倍，且因两函数的底相同，所以在函数 $g$ 中任意赋值 $t=t_0$，在函数 $f$ 中任意赋值 $t=2t_0$ 必定产生同样的值：

$$f(2t_0)=g(t_0)=b^{2t_0}=y_0$$

因此，距离 $y_0J$ 等于 $y_0K$ 的一半，通过类似的推理，对任意 $y$ 值，函数 $g$ 恰在函数 $f$ 与纵轴的正中间。因此可以得出结论：指数**加倍**，会使指数曲线向 $y$ 轴压缩恰好**一半**的距离，而指数减半，将使指数曲线与 $y$ 轴的水平距离**加倍**。

有趣的是两个函数均有相同的纵截距

$$f(0)=g(0)=b^{0}=1$$

指数由 $t$ 变为 $2t$，或变至其它任意倍数，都不会影响纵截距。因

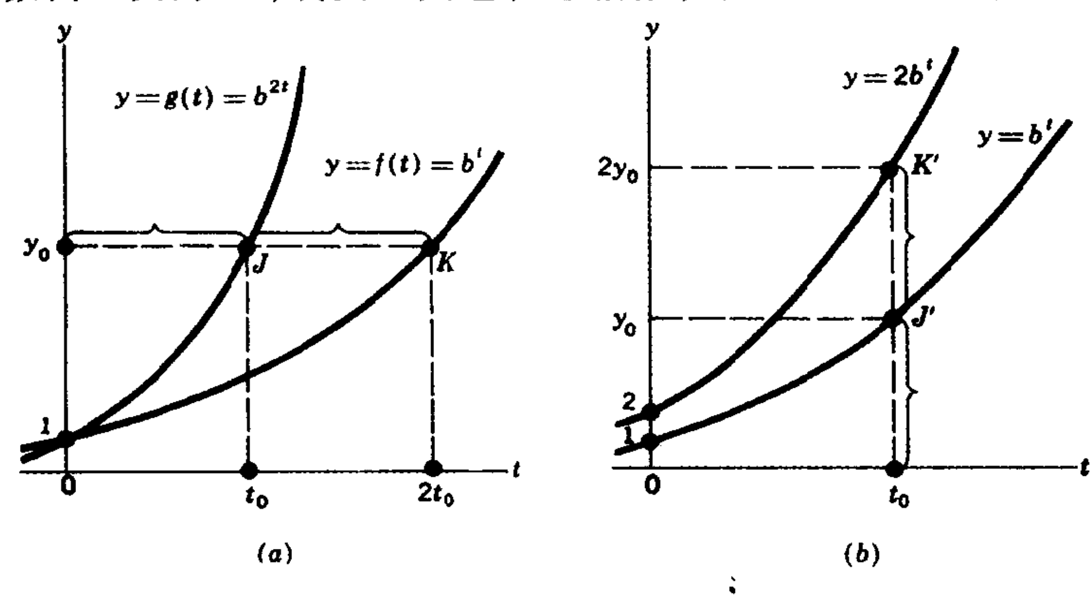



图 10.2

为根据压缩的概念，压缩零的水平距离不会使距离移动。

272 改变指数是变换及推广指数函数(10.1)的一种方式。另一种方式是在 $b^t$ 前加上系数，如 $2b^t$ 等[注意：$2b^t \neq (2b)^t$]。加上这种系数的影响依然是压缩或加大曲线的距离，但这次是使曲线纵向移动。在图 10.2$b$ 中，上面的曲线表示 $y=2b^t$，下面的曲线为 $y=b^t$。对于每个 $t$ 值，前者的高度必定为后者之 2 倍，因为其 $y$ 值恰好为后者之 2 倍。所以，我们有 $t_0J' = J'K'$。还要注意纵截距，它已变成现在这种情形。我们可以得出结论：系数加倍(这里由 1 变至 2)，使曲线与横轴的纵向距离扩大2 倍，系数减半，则使曲线与 $t$ 轴的压缩至原来的一半。

有了上面指数函数两种变换的知识，现在可把指数函数推广为如下形式：

$$y = ab^{ct} \tag{10.2}$$

其中 $a$ 与 $c$ 为“压缩”或“放大”因子。当赋予不同值时，它们将改变指数曲线的位置，从而产生一个指数曲线族。若 $a$ 与 $c$ 为正，则图 10.2 所示之一般图形便适用。但若 $a$ 与 $c$ 均为负，则曲线的图形要作根本的修正(见练习 10.1—5)。

## 一个优先选用的底

273 促使讨论将指数由 $t$ 变至 $ct$ 的问题是换底问题。但即便换底是可行的，为什么要换底呢？理由之一是就数学处理而言，有些底比另一些底更方便。

奇怪的是，在微积分中优先选用的底恰好是一个以符号 $e$ 表示的无理数：

$$e = 2.71828\cdots$$

当在指数函数中运用此指数 $e$ 时，此函数便是**自然指数函数**，比如

$$y=e^{t}\quad y=e^{3t}\quad y=Ae^{rt}$$

这些函数也可以用另一种符号表示

$$y=\exp(t)\quad y=\exp(3t)\quad y=A\exp(rt)$$

其中 exp(exponential 的缩写)表示括号中的式子为 $e$ 的指数。

选择像 $e=2.71828\cdots$ 这样古怪的数作为优先选择的底，当然令人困惑。但这种选择有充分的理由，因为函数 $e^{t}$ 具有这样一种重要性质：它的导数是其自身！即

$$\frac{d}{dt}e^{t}=e^{t}$$

这实际上使微分的工作减少到零。而且借助于这个微分法则——本章后面要对其进行证明——也很容易求出更为复杂的诸如 $y=Ae^{rt}$ 这样的指数函数的导数。为此，首先令 $w=rt$，从而函数变成

$y=Ae^{w}$ 其中 $w=rt$，$A$、$r$ 为常数。则根据链式法则，我们可写出

$$\frac{dy}{dt}=\frac{dy}{dw}\frac{dw}{dt}=Ae^{w}(r)=rAe^{rt}$$

即
$$\frac{d}{dt}Ae^{rt}=rAe^{rt}\tag{10.3}$$

这样底 $e$ 在数学上的方便性便昭然若揭了。

## 练习 10.1

---

1 在一个图中绘出指数函数 $y=3^{t}$ 和 $y=3^{2t}$ 的图形。

(a) 这两个图形是否与图 10.2 a 反映了相同的一般位置关系？

(b) 这两条曲线是否有相同的 $y$ 截距？为什么？

($c$) 在此图中画出函数 $y=3^{3t}$的图形。

274 2 在同一图中绘出指数函数 $y=4^t$ 与 $y=3(4^t)$的图形。

($a$) 两条曲线是否与图 10.2$b$ 表示大致相同的位置关系?

($b$) 两条曲线是否有同样的 $y$ 截距? 为什么?

($c$) 在同一图中绘出函数 $y=\frac{3}{2}(4)^t$ 的图形。

3 认可 $e^t$ 的导数为其自身,运用链式法则求下列函数的 $dy/dt$:

$$(a)y=e^{5t} \quad (b)y=4e^{3t} \quad (c)y=6e^{-2t}$$

4 根据我们对(10.1)的讨论,你能预期函数 $y=e^t$ 是以递增速率单调地递增吗? 通过确定此函数的一阶和二阶导数的符号验证你的答案。在验证答案时,记住此函数的定义域为全体实数之集合,即区间$(-\infty,\infty)$。

5 在(10.2)中,若 $a$ 与 $c$ 被赋予负值,则图 10.2 中的曲线图形便不再适用。通过取($a$) $a=-1$ 代替 $a=1$ 和($b$)$c=-1$ 代替 $c=1$,比较曲线图形的变化。

---

## 10.2 自然指数函数与增长问题

仍未回答的一个相关问题是:数 $e$ 是如何定义的? 除了作为一个方便的底的数学意义外,它是否具有什么经济意义? 自然指数以何种方式应用于经济分析?

### 数 $e$

考察下列函数:

$$f(m)=\left(1+\frac{1}{m}\right)^m \tag{10.4}$$

若 $m$ 被赋予越来越大的值,则 $f(m)$也将取更大的值。具体地,我们求得

$$f(1)=\left(1+\frac{1}{1}\right)^1=2$$

$$f(2)=\left(1+\frac{1}{2}\right)^2=2.25$$

$$f(3)=\left(1+\frac{1}{3}\right)^3=2.37037\cdots$$

$$f(4)=\left(1+\frac{1}{4}\right)^4=2.44141\cdots$$

进而,若 $m$ 增至无穷大,则 $f(m)$ 将收敛于数 $2.71828\cdots\equiv e$;因 275 此,$e$ 可以定义为当 $m\to\infty$ 时,(10.4)的极限:

$$e\equiv\lim_{m\to\infty}f(m)=\lim_{m\to\infty}(1+\frac{1}{m})^m \tag{10.5}$$

$e$ 的近似值为 2.71828 可通过求函数 $\phi(x)$ 的麦克劳林级数来加以验证〔这里使用 $x$ 是为了便于应用展开式(9.14)〕。这样的级数可以给出 $e^x$ 的多项式近似;令多项式中的 $x=1$,则此多项式便可近似表示 $e(=e^1)$的值。如果当级数中的项数增至无穷大时,余项 $R_n$ 趋近于零,即若级数收敛于 $\phi(x)$,则通过使级数中包含的项数足够多,我们便可使 $e$ 值精确至我们所需要的程度。

为此,我们需要此函数的各阶导数。接受 $e^x$ 的一阶导数是 $e^x$ 本身这一事实,我们可以知道,$\phi(x)$的导数便是 $e^x$,类似地,二阶、三阶或者任意高阶导数必定也为 $e^x$。因此,当我们在展开点($x_o=0$)计算所有导数值时,我们得到一个非常简洁的结果:

$$\phi'(0)=\phi''(0)=\cdots=\phi^{(n)}(0)=e^0=1$$

所以,令(9.14)中的 $x_0=0$,$e^x$ 的麦克劳林级数为

$$e^x=\phi(x)=\phi(0)+\phi'(0)x+\frac{\phi''(0)}{2!}x^2+\frac{\phi'''(0)}{3!}x^3+\cdots$$
$$+\frac{\phi^{(n)}(0)}{n!}x^n+R_n$$

$$= 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n + R_n$$

根据(9.15),余项 $R_n$ 可以写成

$$R_n = \frac{\phi^{(n+1)}(p)}{(n+1)!}x^{n+1} = \frac{e^p}{(n+1)!}x^{n+1}$$

$$[\phi^{(n-1)}(x) = e^x; \therefore \phi^{(n+1)}(p) = e^p]$$

当 $n$ 增大时,阶乘 $(n+1)!$ 的值的增加要快于 $x^{n+1}$(对于有限的 $x$)的增加,由此可知当 $n \to \infty$ 时,$R_n \to 0$。所以麦克劳林级数收敛,因而 $e^x$ 的值可以表示成一个无穷级数〔这个表达式含有无穷个相加的项,这些项遵循着一致的可识别的结构模式,而且余项 $R_n$ 在这个表达式中消失了($R_n \to 0$)〕:

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \frac{1}{5!}x^5 + \cdots \quad (10.6)$$

276 作为一个特例,我们求得

$$\begin{aligned} e &= 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \cdots \\ &= 2 + 0.5 + 0.1666667 + 0.0416667 + 0.0083333 \\ &\quad + 0.0013889 + 0.0001984 + 0.0000248 \\ &\quad + 0.0000028 + 0.0000003 + \cdots \\ &= 2.7182819 \end{aligned}$$

因此,若我们希望数字精确到小数点后五位,则我们可以写成 $e = 2.71828$。注意,我们无需为无穷级数中后面各项担心,因为若我们仅关心前五位小数,它们的大小均可忽略不计。

## $e$ 的经济解释

从数学上看,数 $e$ 是(10.5)的极限。但它是否具有某种经济

意义呢？答案是：它可以解释成复利的一种具体计算过程的结果。

假设开始时本金（或资本）为一美元，一个假设的银行家给我们 100% 的不平常的年利率（每年 1 美元利息）。若利息按复利每年计算一次，则年末资产价值为 2 美元；我们以 $V(1)$ 表示此值，其中括号中的数字表示一年内计算复利的次数。

$$V(1) = \text{初始本金}(1 + \text{利息率})$$

$$= 1(1 + 100\%) + \left(1 + \frac{1}{1}\right)^1 = 2$$

但是，若半年计算一次复利，则六个月末利息等于本金的 50%（100% 的一半）。因此在第二个六个月期间，本金为 1.5 美元，此期间利息将按 1.5 美元的 50% 计算。因此年末资产价值为 1.50(1+50%)；即

$$V(2) = (1 + 50\%)(1 + 50\%) = \left(1 + \frac{1}{2}\right)^2$$

通过类似的推理，我们可以写出 $V(3) = \left(1 + \frac{1}{3}\right)^3$，$V(4) = \left(1 + \frac{1}{4}\right)^4$ 等，或者一般地

$$V(m) = \left(1 + \frac{1}{m}\right)^m \tag{10.7}$$

其中 $m$ 表示 1 年内复利计算次数。

在极限情况下，当利息在一年内连续按复利计算时，即 $m$ 取无穷大时，资产价值将以“滚雪球”方式增长，在一年末变成

$$\lim_{m \to \infty} V(m) = \lim_{m \to \infty}\left(1 + \frac{1}{m}\right)^m = e(\text{美元}) \quad [\text{由}(10.5)]$$

因此，如果按年利率 100% 连续计算复利，则 $e = 2.71828$ 可以解释为一美元本金到年末的价值。 277

注意,100%的利息率仅是名义利息率,若一年后一美元变成 $e=2.718$ 美元,则此情况下的实际利率约为每年172%。

## 复利计算与函数 $Ae^{rt}$

刚才讨论的连续复利计算过程可以考虑在如下三个方向一般化:(1)多年复利计算;(2)本金不是一美元;(3)一个不是100%的年名义利息率。

若连续计算复利,1美元本金一年后变成 $e$ 美元,若我们令 $e$ 为第二年的新本金(在第二年,每美元又会增至 $e$ 美元),则第二年末我们的资产显然会成为 $e(e)=e^2$ 美元。同理,第三年末将会变成 $e^3$ 美元,更一般地,$t$ 年末将会成为 $e^t$ 美元。

其次,我们将本金由1美元变成一个不具体的数量 $A$ 美元。这个变化很容易处理:若按100%的名义年利率连续计算复利,$t$ 年后1美元的本金变成 $e^t$ 美元,则可以推断 $A$ 美元将增至 $Ae^t$ 美元。

如果名义利息率不是100%,比如是 $r=0.05(=5\%)$,那将如何呢?这种利率变化的影响是将表达式 $Ae^t$ 变为 $Ae^{rt}$,我们下面还要对此进行证明。初始本金为 $A$ 美元,按名义利息率 $r$ 投资 $t$ 年,复利计算公式必须修正为如下形式:

$$V(m)=A\left(1+\frac{r}{m}\right)^{mt} \tag{10.8}$$

系数 $A$ 的插入,反映了本金由原来的1美元水平变至 $A$ 美元。商 $r/m$ 意味着在一年计算 $m$ 次复利时,每次名义利率 $r$ 只有 $1/m$ 可以应用。最后指数 $mt$ 表示,因利息按复利一年计算 $m$ 次,所以在 $t$ 年将总共计算 $t$ 次。

(10.8)还可以变换成另一种形式

$$V(m) = A\left[\left(1+\frac{r}{m}\right)^{m/r}\right]^{rt} \tag{10.8'}$$

$$= A\left[\left(1+\frac{1}{w}\right)^{w}\right]^{rt} \qquad 其中\ w \equiv \frac{m}{r}$$

随着复利计算次数 $m$ 的增加,新产生的变量 $w$ 必然同时增加,所以,当 $m \to \infty$,我们有 $w \to \infty$,而且由(10.5),(10.8′)括号中的表达式将趋于 $e$。因而,我们以一般化的连续复利计算方法求得的资产价值为

278

$$V \equiv \lim_{m\to\infty} V(m) = Ae^{rt} \tag{10.8''}$$

它与前面的预期是一致的。

注意,在(10.8)中,$t$ 是一个离散(相对于连续的)的变量;它只能取 $1/m$ 的整数倍的值。例如,若 $m=4$(每季度计算一次复利),则 $t$ 仅可以取值 1/4,1/2,3/4,1 等,这表明仅在每个新季度末,$V(m)$才可以取新值。但是,如在(10.8″)中那样,当 $m \to \infty$ 时,$1/m$ 将变得无穷小,而且变量 $t$ 也将变得连续。在此情况下,谈及一年的几分之一,或者令 $t=1.2$ 或 $t=2.35$ 等便合理了。

总之,如在表 10.1 中所概括的那样,表达式 $e$、$e^t$、$Ae^t$ 的经济解释均与连续复利计算有关。

**表 10.1 连续的复利计算**

| 本金<br>美元 | 名义利<br>息率 | 连续计算<br>复利的年份 | 复利计算过程末的<br>资产价值,美元 |
|---|---|---|---|
| 1 | 100%(=1) | 1 | $e$ |
| 1 | 100% | $t$ | $e^t$ |
| $A$ | 100% | $t$ | $Ae^t$ |
| $A$ | $r$ | $t$ | $Ae^{rt}$ |

## 瞬间增长率

应该指出,复利计算是自然指数 $Ae^{rt}$ 性质的一种解释,但不是唯一的解释。复利计算仅仅是指数增长一般过程的一个例子(这里是货币资本总额随时间的增长),我们同样还可以将指数函数应用于人口、财富、有形资本等的增长。

应用于其它内容而非复利计算,则 $Ae^{rt}$ 中的系数 $r$ 便不再表示名义利息率。那么它的经济含义又是什么呢?答案是,$r$ 可以重新解释成函数 $Ae^{rt}$ 的瞬间增长率。〔事实上,这也是为什么我们一开始就选择符号 $r$(rate of growth)的原因。〕已知函数 $V=Ae^{rt}$,它给出每个时点 $t$ 的 $V$ 值,$V$ 的变化率可通过求导得出:

$$\frac{dV}{dt}=rAe^{rt}=rV\ [\text{见}(10.3)]$$

但 $V$ 的增长率也就是以相对值(百分率)表示的 $V$ 的变化率,即可以表示成 $V$ 值自身的比率。因此,对任意给定时点,我们有

279

$$V\text{ 的增长率}\equiv\frac{dV/dt}{V}=\frac{rV}{V}=r \tag{10.9}$$

这与前面的表述是一致的。

关于增长率还有几点要注意。但首先,我们应澄清时间概念的一个基本点:即时点与时期的区别。变量 $V$(表示货币总量、人口数量等)是一个存量概念,它与这样的问题有关:在给定时刻,它存在的数量是多少?因此,$V$ 是一个与时点相关的量;在每一时点,$V$ 取唯一值。而 $V$ 的变化则代表流量,它涉及这样的问题:在给定时段,它产生了多大变化?同样,$V$ 的变化,$V$ 的变化率等必须参照某一特定时期,比如每年来确定。

基于这些理解,我们对(10.9)进行一些说明:

1 (10.9)所定义的增长率是瞬间增长率。因为导数 $dV/dt = rAe^{rt}$在不同的 $t$ 点取不同的值，$V = Ae^{rt}$也同样如此，因此，$dV/dt$ 与 $V$ 的比率仅与特定 $t$ 点（或瞬时点）有关。在这个意义上，此增长率为瞬间增长率。

2 但在目前的情况下，瞬时增长率恰好为常数 $r$，增长率在所有时点保持不变。当然，我们遇到的所有增长情况并非总是如此。

3 尽管增长率 $r$ 是在瞬间，比如在某一特定时点度量的，但其大小仍具有"单位时间（比如每年，若 $t$ 以年为单位）多少百分数"的含义。增长就其本质而言，仅在一定时间内才可能发生。这就是为什么单个静止的图像（记录一个瞬间的状态）永远不能描述（比如说）一个儿童的成长，而两张取自不同时点（比如间隔一年）的相片，则可以描述其成长的原因。

4 对于指数函数 $V = Ae^{rt}$，在所有 $t$ 点，增长的百分率是相同的，但 $V$ 的绝对增量随着时间的增加而增大，这是因为百分率是在越来越大的基数上计算的。

把 $r$ 视为瞬时增长率后，若 $r$ 为常数，则很容易求出形式为 $y = Ae^{rt}$的自然指数函数的增长率。比如，给定函数 $y = 75e^{0.02t}$，我们立即可以读出 $y$ 的增长率为 0.02 或 2%。

## 连续增长与离散增长

280

上面的讨论尽管具有分析意义，但其在经济上的适用性仍有待于讨论，因为实际增长并不总是在连续的基础上发生的，即便复利计算也不是如此。但幸运的是，即使对离散增长的情况，亦即变化是在某一时段中发生而不是每时每刻都发生的情况，运用连续

指数增长函数仍是合理的。

在复利计算频数尽管并非无限，但仍然相对高的情况下，连续增长模式可以视为真实增长模式的一种近似。但更重要的是，我们能够证明离散或不连续增长问题均可以变换成等价的连续形式。

假设我们有如下序列表示的几何增长模式(比如离散的复利计算)：

$$A, A(1+i), A(1+i)^2, A(1+i)^3, \cdots$$

其中 $i$ 表示每期实际利息率，$(1+i)$的指数表示复利计算的时期的数字。如果我们将$(1+i)$视为一个指数表达式中的底 $b$，则上述序列便可以指数函数 $Ab^t$ 来概括，只是由于问题的离散性，$t$ 仅限于取正整数。而且，$b=1+i$ 为正数(即使利率 $i$ 为负，比如 $-0.04$，$b$ 仍为正)，所以它总可以表示为任意大于 1 的任意实数(包括 $e$)的幂。这意味着必然存在一个数 $r$ 使得①

$$1+i=b=e^r$$

因此，我们可将 $Ab^t$ 变换成为自然指数函数：

$$A(1+i)^t=Ab^t=Ae^{rt}$$

对任意给定 $t$ 值(这里，$t$ 取整数值)，函数 $Ae^{rt}$ 当然会同 $A(1+i)^t$ 产生完全相同的值，比如 $A(1+i)=Ae^r$，$A(1+i)^2=Ae^{2r}$ 等。因此，尽管考察的是离散的情况 $A(1+i)^t$，但我们仍可运用连续的自然指数函数 $Ae^{rt}$。这就解释了为什么尽管所有增长模式并不都是连续的，但自然指数函数在经济分析中仍然得到广泛应用的原因。

① 给定具体 $b$ 值，求数 $r$ 的方法将在 10.4 节讨论。

## 贴现与负增长

现在我们从复利计算转至与其有密切关系的概念——贴现。在复利问题中,我们要由一个给定现值 $A$(初始本金),计算未来值 $V$(本金加利息)。而贴现问题则与其相反:由一个已知的 $t$ 年后可利用的总额 $V$,求现值 $A$。 281

首先我们考察离散的情况。若按年利息率 $i$ 每年计算复利至 $t$ 年后,本金 $A$ 的数量增至未来值 $A(1+i)^t$,即如果

$$V = A(1+i)^t$$

则通过将方程两边以非零表达式 $(1+i)^t$ 除,我们可以得到贴现公式:

$$A = \frac{V}{(1+i)^t} = V(1+i)^{-t} \tag{10.10}$$

此式包含一个负指数。应认识到,在此式中,$V$ 与 $A$ 的角色已颠倒了:$V$ 是已知值,而 $A$ 则是有待于通过 $i$(贴现率)、$t$(年数)和 $V$ 来计算的未知值。

类似地,在连续的情况下,若按利率 $r$ 连续计算复利至 $t$ 年后,本金 $A$ 增至 $Ae^{rt}$,即按公式

$$V = Ae^{rt}$$

则通过以 $e^{rt}$ 除方程两边,我们便可得到相应的连续贴现公式

$$A = \frac{V}{e^{rt}} = Ve^{-rt} \tag{10.11}$$

同样,这里 $A$(而非 $V$)是有待于通过给定未来值 $V$、名义贴现率 $r$、年份数 $t$ 来计算的未知值。

将(10.11)看作指数增长函数,我们立即可以读出 $-r$ 是 $A$ 的

瞬时增长率。作为负值,它有时也被称作缩减率(*rate of decay*)。正如计算复利说明了增长过程一样,贴现描述了负的增长。

**练习 10.2**

---

1 运用(10.6)中 $e^x$ 的无穷级数形式,求下列函数的近似值:

$(a)e^2$ $(b)\sqrt{e}(=e^{\frac{1}{2}})$

(每项计算至小数点后面第 3 位,再四舍五入,计算级数各项至你得到 0.000 的项为止。)

2 给定函数 $\phi(x)=e^{2x}$:

(*a*) 写出其麦克劳林级数的多项式部分 $P_n$。

(*b*) 写出余项 $R_n$ 的拉格朗日型。确定当 $n\to\infty$ 时,是否有 $R_n\to 0$,即确定级数是否收敛于 $\phi(x)$。

282 (*c*) 若级数收敛,则 $\phi(x)$ 可以用一无穷级数表示,写出此级数。

3 写出下列值的指数表达式:

(*a*) 本金 10 美元,按利率 5%连续计算复利 3 年。

(*b*) 本金 690 美元,按利率 4%连续计算复利 2 年。

(这些利率均为年名义利息率。)

4 在下列各式中,*y* 的瞬时增长率为多少?

$(a)y=e^{0.07t}$ $(c)y=Ae^{0.2t}$

$(b)y=12e^{0.03t}$ $(d)y=0.03e^t$

5 证明:函数 $y_1=Ae^{rt}$(计算复利)与 $y_2=Ae^{-rt}$(贴现)关于 *y* 轴镜像对称〔参见练习 10.1-5 (*b*)部分〕。

---

## 10.3 对数

指数函数与对数函数(简写为 log 函数)有着密切的关系。在

我们讨论 log 函数之前，我们必须先理解对数的含义。

### 对数的含义

若我们有两个数 4 和 16，二者可以通过方程 $4^2=16$ 联系起来，则我们可以定义指数 2 为以 4 为底 16 的对数，并写成：

$$\log_4{}^{16}=2$$

由此例应当清楚，对数只不过是一个幂，底(4)自乘到这个幂便得到一个特定值(16)。一般而言，我们可以表述如下：

$$y=b^t \quad \Leftrightarrow \quad t=\log_b y \tag{10.12}$$

它表明，以 $b$ 为底 $y$ 的对数(表示为 $\log_b y$)等于底 $b$ 自乘以得到 $y$ 的幂。因此，写出下式

$$b^{\log_b y}=y$$

是正确的，尽管有些重复。

在讨论指数函数时，我们强调函数 $y=b^t(b>1)$ 是单调递增的。这意味着对任意正值 $y$，存在一个唯一的指数 $t$(不必一定为正)使得 $y=b^t$；而且，如在图 10.2 中可以见到那样，$y$ 值越大，$t$ 必定越大。用对数来表达这个含义，即指数函数的单调性意味着任意正数 $y$ 必定有底 $b>1$ 的唯一的对数 $t$，使得 $y$ 越大，其对数越大。如图 10.1 和 10.2 所示，在指数函数 $y=b^t$ 中，$y$ 必然为正，所以负数或零没有对数。 283

### 常用对数与自然对数

对数的底 $b>1$，并不必限定为某个特定的数。但在实际应用中，有两个数，即 10 和 $e$，被广泛地选用为底。当 10 作为底时，此对数被称作常用对数，以 $\log_{10}$(若文中含义清楚，亦可简写为 log)

表示。另一方面，以 $e$ 为底的对数被称作自然对数，或者以 $\log_e$ 表示，或者以 ln（对自然对数）表示。若在具体的上下文中不至造成歧义，也可以使用符号 log（舍去下标 $e$）。

在计算工作中常用的常用对数，举例如下：

$$\log_{10} 1000 = 3 \qquad [\text{因为 } 10^3 = 1000]$$

$$\log_{10} 100 = 2 \qquad [\text{因为 } 10^2 = 100]$$

$$\log_{10} 10 = 1 \qquad [\text{因为 } 10^1 = 10]$$

$$\log_{10} 1 = 0 \qquad [\text{因为 } 10^0 = 1]$$

$$\log_{10} 0.1 = -1 \qquad [\text{因为 } 10^{-1} = 0.1]$$

$$\log_{10} 0.01 = -2 \qquad [\text{因为 } 10^{-2} = 0.01]$$

观察紧邻等号左边的数集与紧邻等号右边的数集的关系。由此可知，10 与 100 之间的数的常用对数必在 1 与 2 之间；1 与 10 之间的常用对数必定为正分数，等等。确切的对数值可通过常用对数表，或具有对数功能的电子计算器得到。①

但在分析工作中，使用自然对数远比使用常用对数更为方便。因为根据对数定义，我们有这种关系：

$$y = e^t \quad \Leftrightarrow \quad t = \log_e y \text{（或 } t = \ln y\text{）} \qquad (10.13)$$

很容易看到，指数函数中 $e$ 在分析中的方便性自然会转移到以 $e$ 为底的对数中去。

下面的例子用于说明自然对数：

284

$$\ln e^3 = \log_e e^3 = 3$$

① 更基本地，对数值，像 $e$ 值一样，可以借助于对数函数的麦克劳林级数求得，方法与（10.6）式描述的方法类似。但我们在这里不触及这个问题。

$$\ln e^2 = \log_e e^2 = 2$$

$$\ln e^1 = \log_e e^1 = 1$$

$$\ln 1 = \log_e e^0 = 0$$

$$\ln \frac{1}{e} = \log_e e^{-1} = -1$$

上述例子表明的一般原理是:给定表达式 $e^n$,其中 $n$ 为任意实数,我们可立即读出指数 $n$ 为 $e^n$ 的自然对数。因此,一般地我们有结果 $\ln e^n = n$。[①]

常用对数与自然对数可互相转换,即对数的底像指数式的底一样,均可改换。在学完对数的基本法则后,我们将给出互换公式。

## 对数法则

对数与指数类似,因此,对数也遵守与在 2.5 节中介绍的指数法则密切相关的一些法则。这些法则对简化数学计算具有极大益处。前三个法则仅按自然对数表述,当以 $\log_b$ 取代 ln 时,它们依然成立。

$$\ln(uv) = \ln u + \ln v \qquad (u, v > 0)$$

**法则 I (积的对数法则)**

**例 1** $\ln(e^6 e^4) = \ln e^6 + \ln e^4 = 6 + 4 = 10$

**例 2** $\ln(Ae^7) = \ln A + \ln e^7 = \ln A + 7$

---

① 作为一种记忆方法,我们可以观察到,当符号 ln(或 $\log_e$)置于 $e^n$ 左边时,符号 ln 似乎消掉了符号 $e$,剩下答案 $n$。

证明：由定义，$\ln u$ 是 $e$ 的自乘的幂以得到 $u$，从而 $e^{\ln u}=u$。[①] 类似地，我们有 $e^{\ln v}=v$，以及 $e^{\ln(uv)}=uv$。后者是 $uv$ 的指数式。但是 $uv$ 的另一个表达式可由 $u$ 与 $v$ 直接相乘得到：

$$uv=e^{\ln u}e^{\ln v}=e^{\ln u+\ln v}$$

因此，使 $uv$ 的两个表达式相等，我们有

$$e^{\ln(uv)}=e^{\ln u+\ln v} \quad \text{或} \quad \ln(uv)=\ln u+\ln v$$

285 **法则 II** （**商的对数**）$\ln(u/v)=\ln u-\ln v \qquad (u, v>0)$

**例 3** $\ln(e^2/c)=\ln e^2-\ln c=2-\ln c$

**例 4** $\ln(e^2/e^5)=\ln e^2-\ln e^5=2-5=-3$

此法则的证明与法则 I 极为类似，所以留给读者作为练习，自行证明。

**法则 III** （**幂的对数**）$\ln u^a=a\ln u \qquad (u>0)$

**例 5** $\ln e^{15}=15\ln e=15$

**例 6** $\ln A^3=3\ln A$

证明：由定义，$e^{\ln u}=u$；类似地，$e^{\ln u^a}=u^a$。但是，$u^a$ 的另一表达式还可以这样建立：

$$u^a=(e^{\ln u})^a=e^{a\ln u}$$

通过使 $u^a$ 的两个表达式相等，我们得到所需要的结果：$\ln u^a=a\ln u$。

这三个法则对简化某些类型问题的数学计算是非常有用的工具。法则 I 通过对数，将乘积运算（$uv$）转换成加和运算（$\ln u+\ln v$）；法则 II 把除法（$u/v$）转换为减法（$\ln u-\ln v$）；而法则 III 则使我们将幂简化为乘积常数。而且这些法则还可以组合运用。

---

① 注意，当 $e$ 被自乘至幂 $\ln u$ 时，符号 $e$ 与符号 ln 似乎又相互抵销，剩下答案 $u$。

**例 7**　$\ln(uv^{a}) = \ln u + \ln v^{a} = \ln u + a \ln v$

但读者应注意，若我们首先有加和表达式，则对数对简化计算毫无帮助。特别是，应当记住

$\ln(u \pm v) \neq \ln u \pm \ln v$

现在我们引入关于对数底的变化的另外两个法则。

**法则 IV**　**（对数底的转换）**　$\log_b u = (\log_b e)(\log_e u) \qquad (u > 0)$

此法则与链式法则有神似之处（"链"$b \nearrow^{e} \searrow e \nearrow^{u}$ 便是证明），它使我们由对数 $\log_b u$（以 $b$ 为底）推导出 $\log_e u$（以 $e$ 为底），反之亦然。

证明：令 $u = e^{p}$，从而 $p = \log_e u$。则由此可知

$$\log_b u = \log_b e^{p} = p \log_b e = (\log_e u)(\log_b e)$$

法则 IV 可以很容易推广为

$$\log_b u = (\log_b c)(\log_c u)$$

其中 $c$ 是不同于 $b$ 的底。

286

**法则 V**　**（对数底的转换）**　$\log_b e = \dfrac{1}{\log_e b}$

此法则类似于反函数的微分法则，它使我们在知道以 $b$ 为底 $e$ 的对数的基础上，马上就可以得到以 $e$ 为底 $b$ 的对数。反之亦然。（此法则也可以推广成这种形式：$\log_b c = 1/\log_c b$）

证明：作为法则 IV 的一个应用，令 $u = b$，则我们有

$$\log_b b = (\log_b e)(\log_e b)$$

但表达式左侧是 $\log_b b = 1$，因此，$\log_b e$ 与 $\log_e b$ 必定互为倒数，这正是法则 V 的结论。

由上面后两个法则，很容易推导出常用对数与自然对数间的转换公式：

$$\log_{10} N = (\log_{10} e)(\log_e N) = 0.4343 \log_e N$$

$$\log_e N = (\log_e 10)(\log_{10} N) = 2.3026 \log_{10} N \quad (10.14)$$

至于 $N$,它是一个正实数。每个公式中第一个等号很容易由法则 IV 得到证明。在第一个公式中,值 0.4343(2.71828 的常用对数)可由常用对数表或用计算器求出。在第二个公式中,值 2.3026(10 的自然对数)仅是 0.4343 的倒数,这样计算是由于法则 V。

**例 8** $\log_e 100 = 2.3026(\log_{10} 100) = 2.3026(2) = 4.6052$。反之,我们有 $\log_{10} 100 = 0.4343(\log_e 100) = 0.4343(4.6052) = 2$。

**一个应用**

287 上述对数法则使我们可以解一些简单的指数方程(令指数函数等于零)。例如,若我们求满足如下方程的 $x$ 的值

$$ab^x - c = 0 \qquad (a, b, c > 0)$$

首先,通过使用对数,我们将此指数方程转换成线性方程,然后再解它。为此,应首先将 $c$ 项移至左边:

$$ab^x = c$$

虽然对和 $(ab^x - c)$ 并无简单的对数表达式,但对乘积 $ab^x$ 和单项 $c$ 则可方便地取对数。因此,将 $c$ 移项后,对方程两边取对数(比如以 10 为底),我们有

$$\log a + x \log b = \log c$$

这是关于变量 $x$ 的线性方程,其解为

$$x = \frac{\log c - \log a}{\log b}$$

## 练习 10.3

1 下列对数的值为多少?

(a) $\log_{10}10,000$　　(c) $\log_3 81$

(b) $\log_{10}0.0001$　　(d) $\log_5 3125$

2 计算下列对数:

(a) $\ln e^2$　　(c) $\ln(1/e^3)$　　(e) $(e^{\ln 3})!$

(b) $\log_e e^{-4}$　　(d) $\log_e(1/e^2)$　　(f) $\ln e^x - e^{\ln x}$

3 应用对数法则计算下列对数:

(a) $\log_{10}(100)^{14}$　　(d) $\ln Ae^2$

(b) $\log_{10}\frac{1}{100}$　　(e) $\ln ABe^{-4}$

(c) $\ln(3/B)$　　(f) $(\log_4 e)(\log_e 64)$

4 下列哪些等式成立?

(a) $\ln u - 2 = \ln\frac{u}{e^2}$　　(c) $\ln u + \ln v - \ln w = \ln\frac{uv}{w}$

(b) $3 + \ln v = \ln\frac{e^3}{v}$　　(d) $\ln 3 + \ln 5 = \ln 8$

5 证明 $\ln(u/v) = \ln u - \ln v$。

## 10.4 对数函数

当一个变量被表示成另一个变量的对数的函数时,此函数被称为**对数函数**。在(10.12)和(10.13)式中,我们已看到两种类型的对数函数,即

$$t = \log_b y \quad 和 \quad t = \log_e y\ (= \ln y)$$

二者的差别仅在于对数的底不同。

如前所述，对数函数是某些指数函数的反函数。考察上述两个对数函数可以进一步确认，它们确实分别是两个指数函数的反函数：

$$y = b^t \quad 和 \quad y = e^t$$

因为所列对数函数是对应的指数函数的因变量与自变量角色互换的结果。当然，读者应当认识到，这里把符号 $t$ 用作一般的符号，它并不一定表示时间。即使当它代表时间时，它以因变量出现并不意味着时间是由某个变量 $y$ 决定的，它仅意味着一个给定 $y$ 值是与一个唯一的时点相联系的。

作为单调递增（指数）函数的反函数，对数函数也是单调递增的，这与我们前面的描述是一致的：对任意给定的底，数越大，其对数也越大。这个性质用符号可按照下面两个命题来表示：对于两个 $y$ 的正值（$y_1$ 和 $y_2$）

$$\begin{aligned} \ln y_1 = \ln y_2 \quad &\Leftrightarrow \quad y_1 = y_2 \\ \ln y_1 > \ln y_2 \quad &\Leftrightarrow \quad y_1 > y_2 \end{aligned} \tag{10.15}$$

当然，若我们以 $\log_b$ 代替 ln，这两个命题仍成立。

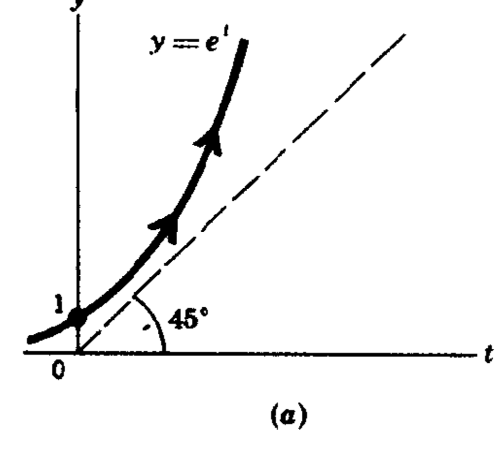
## 图形形状

对数函数的单调性及其它一般性质可由其图形清楚地观察到。给定指数函数 $y = e^t$ 的图形，我们互换两轴的位置，并重画原来的图形，便可得到对应的对数函数的图形。重画的图形即图 10.3。注意，若将图 $b$ 与图 $a$ 重叠，$x$ 轴对 $x$ 轴，$y$ 轴对 $y$ 轴，则两条曲线将完全重合。另一方面，如在图 10.3 所示（互换两轴），则

两条曲线对通过原点的 45°线呈镜像对称(任何一对反函数的图形都是对称的)。

这种镜像对称关系有几点值得注意的含义。第一,尽管两条曲线均为单调递增的,但对数曲线以递减的速率递增,而指数函数则与其相反,以递增的速率递增。另一个有趣的对比是,指数函数的值域为正,而对数函数的定义域为正。(当然,对对数函数定义域的限制仅仅是只有正数方可有对数的另一种表述。)镜像关系的第三个特点是,正如 $y=e^t$ 具有纵截距 1 一样,对数函数 $t=\log_e y$ 必定在 $y=1$ 处通过横轴,表明 $\log_e 1=0$。由于横截距不受对数的 289 底影响(比如,$\log_{10}1=0$),所以由一般图形图 10.3$b$,我们可以推论

$$\left.\begin{array}{r}0<y<1\\ y=1\\ y>1\end{array}\right\}\Leftrightarrow\left\{\begin{array}{l}\log y<0\\ \log y=0\\ \log y>0\end{array}\right. \qquad (10.16)$$



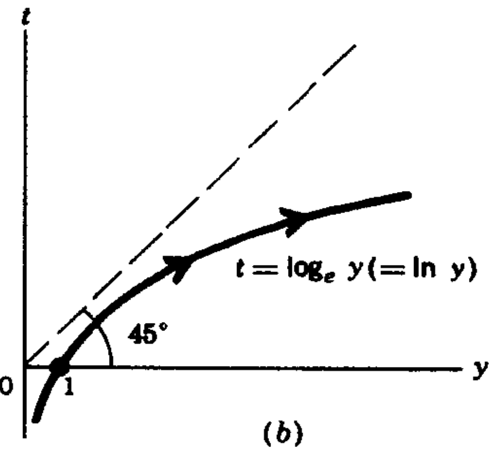




图 10.3

为对此进行验证,读者可检验 10.3 节中给出的两组常用对数与自然对数的例子。此外,还要注意:

$$\log y \to \begin{Bmatrix} \infty \\ -\infty \end{Bmatrix} \quad 当\ y \to \begin{cases} \infty \\ 0^{+} \end{cases} \qquad (10.16')$$

对数函数与指数函数在图 10.3 中的图形比较是以简单对数函数 $y = e^t$ 和 $t = \ln y$ 为基础的。若我们将一般化的指数函数 $y = Ae^{rt}$ 与对应的对数函数相比较,将会得到同样的结果。由于(正)常数 $A$ 和 $r$ 表示压缩或放大指数曲线,$y = Ae^{rt}$ 的图形与图 10.3$a$ 中的一般形状仍是类似的,只是纵截距不再是 $y = 1$,而是 $y = A$。(当 $t = 0$ 时,我们有 $y = Ae^0 = A$。)相应地,其反函数的横截距也是 $y = A$。一般而言,相应的对数曲线与指数曲线关于 45°线呈镜像关系。

如果要求出 $y = Ae^{rt}$ 的反函数的具体表达式,则对指数函数的两边取自然对数〔根据命题(10.15),取对数后方程仍相等〕,然后解出 $t$ 即可:

$$\ln y = \ln\,(Ae^{rt}) = \ln A + rt \ln e = \ln A + rt$$

因此

$$t = \frac{\ln y - \ln A}{r} \qquad (r \neq 0) \qquad (10.17)$$

290 这个结果,即对数函数,便是指数函数 $y = Ae^{rt}$ 的反函数。如前所述,函数(10.17)的横截距为 $y = A$。因为当 $y = A$ 时,我们有 $\ln y = \ln A$,因此 $t = 0$。

## 换　底

在 10.2 节中曾表明,指数函数 $y = Ab^t$ 总可以转换成为自然指数函数 $y = Ae^{rt}$。现在我们准备推导出一个转换公式。但我们要考察的是如何将更一般的表达式 $Ab^{ct}$ 转换为 $Ae^{rt}$,而不是转换

$Ab^t$。因为问题的实质是由给定的 $b$、$c$ 求 $r$，使得

$$e^r = b^c$$

所要做的就是将 $r$ 表示成 $b$ 和 $c$ 的函数。对上面的方程两边取自然对数，便可以很轻松地完成这个工作：

$$\ln e^r = \ln b^c$$

方程的左边显然等于 $r$，所以所求函数(转换公式)为

$$r = \ln b^c = c\ \ln b \tag{10.18}$$

这表明，函数 $y = Ab^{ct}$ 总可以改写成自然底的形式：

$$y = Ae^{(c\ln b)t}$$

**例 1** 将 $y = 2^t$ 变换成自然指数函数。这里，我们有 $A = 1$，$b = 2$ 及 $c = 1$。因此 $r = c\ln b = \ln 2$，且所求指数函数为

$$y = Ae^{rt} = e^{(\ln 2)t}$$

若愿意，我们还可以运用(10.14)和常用对数表，计算出(ln2)的数值：

$$\ln 2 = 2.3026\log_{10}2 = 2.3026(0.3010) = 0.6931 \tag{10.19}$$

则我们还可以将前面的结果写成 $y = e^{0.6931t}$。

**例 2** 将 $y = 3(5)^{2t}$ 转换为自然指数函数。在本例中，$A = 3$，$b = 5$，$c = 2$，由公式(10.18)可知 $r = 2\ln 5$。因此，所求方程为

$$y = Ae^{rt} = 3e^{(2\ln 5)t}$$

同样，若愿意，我们还可以计算出

$$2\ln 5 = \ln 25 = 2.3026\ \log_{10}25 = 2.3026(1.3979) = 3.2188$$

所以前面的结果还可以表示成 $y = 3e^{3.2188t}$。

当然，将形式为 $t = \log_b y$ 的对数函数转换为等价的自然对数函数也是可以的。为此，只需应用对数法则 IV 便足矣。此法则

可表示为

$$\log_b y = (\log_b e)(\log_e y)$$

将此结果直接带入给定对数函数，马上便可得到所求的自然对数函数：

$$\begin{aligned} t &= \log_b y = (\log_b e)(\log_e y) \\ &= \frac{1}{\log_e b}\log_e y \qquad [\text{由对数法则 V}] \\ &= \frac{\ln y}{\ln b} \end{aligned}$$

运用同样的方法，我们还可将更一般的对数函数 $t = a \log_b(cy)$ 变换为等价形式

$$t = a(\log_b e)(\log_e c y) = \frac{a}{\log_e b}\log_e(cy) = \frac{a}{\ln b}\ln(cy)$$

**例 3** 将函数 $t = \log_2 y$ 变换为自然对数形式。因为在本例中 $b = 2, a = c = 1$，所以所求结果为

$$t = \frac{1}{\ln 2}\ln y$$

但根据(10.19)，我们也可将其表示成 $t = (1/0.6931)\ln y$。

**例 4** 将函数 $t = 7 \log_{10} 2y$ 变换成自然对数。本例中的常数为 $a = 7, b = 10, c = 2$，因而所求函数为

$$t = \frac{7}{\ln 10}\ln 2y$$

如(10.14)所示，因 $\ln 10 = 2.3026$，所以上面的函数可以重写为 $t = (7/2.3026)\ln 2y = 3.0400 \ln 2y$。

在上面的讨论中，当函数为对数函数时，我们一直遵循将 $t$ 表示成 $y$ 的做法。我们这样做的唯一原因是我们希望强调指数函

数与对数函数间的反函数关系。当我们仅研究对数函数时，我们可以按惯例写成 $y=\ln t$（而非 $t=\ln y$）。当然，符号上的这种变换对分析不产生任何影响。

**练习 10.4**

---

1 $y=Ae^{rt}$的反函数(10.17)要求 $r$ 为非零。从原指数函数角度看，这个要求意味着什么？

2 (a)画出指数函数 $y=Ae^{rt}$的图形。指出纵截距的值。 292

(b)画出对数函数 $t=\dfrac{\ln y-\ln A}{r}$的图形，指出其横截距的值。

3 求 $y=ab^{ct}$的反函数。

4 将下列函数转换成自然指数形式：

(a) $y=8^{3t}$　　(c) $y=5(5)^t$

(b) $y=2(7)^{2t}$　　(d) $y=2(15)^{4t}$

5 将下列函数转换成自然对数形式：

(a) $t=\log_7 y$　　(c) $t=3\log_{15}9y$

(b) $t=\log_8 3y$　　(d) $t=2\log_{10}y$

6 已知下列离散复利率($i$)，求与之等价的连续复利年名义利率($r$)：

(a) 年利率5%，按年计算复利。

(b) 年利率5%，按半年计算复利。

(c) 年利率6%，按半年计算复利。

(d) 年利率6%，按季度计算复利。

---

## 10.5　指数函数与对数函数的导数

前面曾指出，函数 $e^t$ 的导数为其自身。实际上，自然对数函数 $\ln t$ 也具有非常方便的导数，即 $d(\ln t)/dt=1/t$。这一事实强

化了我们对底 $e$ 的偏爱。现在我们来证明这两个导数公式的正确性，然后再推导指数函数和对数函数的导数公式的某些变式。

### 对数函数求导法则

对数函数 $y=\ln t$ 的导数为

$$\frac{d}{dt}\ln t=\frac{1}{t}$$

为证明此法则，我们回顾一下，根据定义，函数 $y=f(t)=\ln t$ 的导数在 $t=N$ 处具有如下值

$$f(N)=\lim_{t\to N}\frac{f(t)-f(N)}{t-N}=\lim_{t\to N}\frac{\ln t-\ln N}{t-N}=\lim_{t\to N}\frac{\ln(t/N)}{t-N}$$

[由对数法则 II]

293 现在我们引入简写符号 $m=\frac{N}{t-N}$，则我们可以写出 $\frac{1}{t-N}=\frac{m}{N}$，以及 $\frac{t}{N}=1+\frac{t-N}{N}=1+\frac{1}{m}$，因此，上面极限符号右边的表达式可变换成如下形式：

$$\frac{1}{t-N}\ln\frac{t}{N}=\frac{m}{N}\ln\left(1+\frac{1}{m}\right)=\frac{1}{N}\ln\left(1+\frac{1}{m}\right)^m$$

[根据对数法则 III]

注意，当 $t\to N$ 时，$m$ 将趋于无穷大，所以要求出所需求的导数值，我们可以取当 $m\to\infty$ 时最后一个表达式的极限：

$$f'(N)=\lim_{m\to\infty}\frac{1}{N}\ln\left(1+\frac{1}{m}\right)^m=\frac{1}{N}\ln e=\frac{1}{N}$$

[由(10.5)]

但是，因 $N$ 可以为使对数有定义的任意数，所以我们可以将此结果一般化，并写成 $f'(t)=d(\ln t)/dt=1/t$。对数求导法则得证。

### 指数函数求导法则

函数 $y=e^t$ 的导数为

$$\frac{d}{dt}e^t=e^t$$

由对数函数求导法则很容易得出这一结果。我们知道函数 $y=e^t$ 的反函数为 $t=\ln y$,后者的导数为 $dt/dy=1/y$。因此由反函数法则,我们立即可得出

$$\frac{d}{dt}e^t=\frac{dy}{dt}=\frac{1}{dt/dy}=\frac{1}{1/y}=y=e^t$$

### 推广的指数函数求导法则

上述两个法则可推广至表达式 $e^t$ 与 $\ln t$ 中的变量 $t$ 可为 $t$ 的某些函数,比如 $f(t)$代替的情况。两个法则的一般化形式为:

$$\frac{d}{dt}e^{f(t)}=f'(t)e^{f(t)} \qquad \left[\text{或}\frac{d}{dt}e^u=e^u\frac{du}{dt}\right]$$

$$\frac{d}{dt}\ln f(t)=\frac{f'(t)}{f(t)} \qquad \left[\text{或}\frac{d}{dt}\ln v=\frac{1}{v}\frac{dv}{dt}\right] \tag{10.20}$$

(10.20)通过直接应用链式法则便可得到证明。已知函数 $y=e^{f(t)}$,我们可首先令 $u=f(t)$,从而 $y=e^u$。则由链式法则,导数为

$$\frac{d}{dt}e^{f(t)}=\frac{d}{dt}e^u=\frac{d}{du}e^u\frac{du}{dt}=e^u\frac{du}{dt}=e^{f(t)}f'(t)$$

类似地,已知函数 $y=\ln f(t)$,我们可以首先令 $v=f(t)$,从 294 而形成链;$y=\ln v$,其中 $v=f(t)$,则由链式法则,我们有

$$\frac{d}{dt}\ln f(t)=\frac{d}{dt}\ln v=\frac{d}{dv}\ln v\frac{dv}{dt}=\frac{1}{v}\frac{dv}{dt}=\frac{1}{f(t)}f'(t)$$

注意,(10.20)与较简单的法则 $de^t/dt=e^t$ 和 $d(\ln t)/dt=1/t$ 相比较,真正的修正只是引入了一个乘数因子 $f'(t)$。

**例 1** 求函数 $y=e^{rt}$ 的导数。这里指数为 $rt=f(t)$,有 $f'(t)=r$,因此

$$\frac{dy}{dt}=\frac{d}{dt}e^{rt}=re^{rt}$$

**例 2** 由函数 $y=e^{-t}$ 求导数 $dy/dt$。在本例中,$f(t)=-t$,所以 $f'(t)=-1$,因而

$$\frac{dy}{dt}=\frac{d}{dt}e^{-t}=-e^{-t}$$

**例 3** 由函数 $y=\ln at$ 求 $dy/dt$。因在本例中 $f(t)=at$,$f'(t)=a$,所以导数为

$$\frac{d}{dt}\ln at=\frac{a}{at}=\frac{1}{t}$$

非常有趣的是,它与 $y=\ln t$ 的导数是一致的。

此例说明了这样一个事实:在对数表达式内与 $t$ 相乘的常数在求导过程中被消掉了。但要注意,对于常数 $k$,我们有

$$\frac{d}{dt}k\ln t=k\frac{d}{dt}\ln t=\frac{k}{t}$$

所以在对数表达式外与之相乘的常数在求导过程中仍得到保留。

**例 4** 求函数 $y=\ln t^c$ 的导数,因 $f(t)=t^c$,$f'(t)=ct^{c-1}$,所以由(10.20)得到

$$\frac{d}{dt}\ln t^c=\frac{ct^{c-1}}{t^c}=\frac{c}{t}$$

**例 5** 由 $y=t^3\ln t^2$ 求 $dy/dt$。因为此函数是 $t^3$ 与 $\ln t^2$ 的

积，所以需应用积的求导法则：

$$\begin{aligned}\frac{dy}{dt} &= t^3\frac{d}{dt}\ln t^2 + \ln t^2\frac{d}{dt}t^3 \\ &= t^3\left(\frac{2t}{t^2}\right) + (\ln t^2)(3t^2) \\ &= 2t^2 + 3t^2(2\ln t) \qquad [\text{对数求导法则(III)}] \\ &= 2t^2(1 + 3\ln t)\end{aligned}$$

## 底 *b* 的情况

对于以 $b$ 为底的指数函数与对数函数，导数为 295

$$\frac{d}{dt}b^t = b^t\ln b \quad \left[\text{注意：}\frac{d}{dt}b^t \neq tb^{t-1}\right]$$

$$\frac{d}{dt}\log_b t = \frac{1}{t\ln b} \qquad (10.21)$$

注意在底为 $e$ 的特定情况下（$b=e$ 时），我们有 $\ln b = \ln e = 1$，所以这两个导数将分别简化为 $(d/dt)e^t = e^t$ 和 $(d/dt)\ln t = 1/t$。

(10.21) 的证明并不难。对于 $b^t$ 的情况，根据恒等式 $b \equiv e^{\ln b}$，我们可以写出

$$b^t = e^{(\ln b)t} = e^{t\ln b}$$

（我们写成 $t\ln b$，而不写成 $\ln bt$，以强调 $t$ 不是对数表达式中的一部分。）因此，

$$\begin{aligned}\frac{d}{dt}b^t &= \frac{d}{dt}e^{t\ln b} = (\ln b)(e^{t\ln b}) \qquad [\text{由(10.20)}] \\ &= (\ln b)(b^t) = b^t\ln b\end{aligned}$$

要证明(10.21)的第二部分，我们需依靠对数的基本性质

$$\log_b t = (\log_b e)(\log_e t) = \frac{1}{\ln b}\ln t$$

它使我们得到导数

$$\frac{d}{dt}\log_b t = \frac{d}{dt}\left(\frac{1}{\ln b}\ln t\right) = \frac{1}{\ln b}\frac{d}{dt}\ln t = \frac{1}{\ln b}\left(\frac{1}{t}\right)$$

这两个公式的更一般的形式为

$$\frac{d}{dt}b^{f(t)} = f'(t)b^{f(t)}\ln b$$

$$\frac{d}{dt}\log_b f(t) = \frac{f'(t)}{f(t)}\frac{1}{\ln b} \qquad (10.21')$$

同样，若 $b=e$，则 $\ln b=1$，则此式可以简化为(10.20)。

**例6** 求函数 $y=12^{1-t}$ 的导数。这里，$b=12$，$f(t)=1-t$，$f'(t)=-1$，所以

$$\frac{dy}{dt} = -(12)^{1-t}\ln 12$$

296

## 高阶导数

同其它类型的函数一样，指数函数与对数函数的高阶导数仅仅是重复求导的结果。

**例7** 求 $y=b^t(b>1)$ 的二阶导数。由(10.21)知，一阶导数为 $y'(t)=b^t\ln b$（当然，其中 $\ln b$ 为常数），于是对 $t$ 再微分，我们有

$$y''(t) = \frac{d}{dt}y'(t) = \left(\frac{d}{dt}b^t\right)\ln b = (b^t\ln b)\ln b = b^t(\ln b)^2$$

注意，$y=b^t$ 总为正，且因 $b>1$，所以 $\ln b$ 也总为正[由(10.16)]，因而 $y'(t)=b^t\ln b$ 必然也为正。这一事实支持了我们前面的关于指数函数 $y=b^t$ 以递增的速率单调递增的观点。

**例 8**　求 $y=\ln t$ 的二阶导数。其一阶导数为 $y'=1/t=t^{-1}$,因此二阶导数为

$$y''=-t^{-2}=-\frac{1}{t^2}$$

由于此函数的定义域为开区间$(0,\infty)$,所以 $y'=1/t$ 必为正。而 $y''$必为负。这两个结论的结合则支持了我们前面关于对数函数 $y=\ln t$ 以递减速率单调递增的判断。

## 一个应用

对数函数的主要优点之一是它能将乘法转化为加法,将除法转化为减法,当我们对任意类型的函数的积或商微分时,我们就可以利用这一性质。

**例 9**　求函数

$$y=\frac{x^2}{(x+3)(2x+1)}$$

的导数 $dy/dx$。我们首先对方程两边取自然对数,而不是应用积和商的求导法则,从而得到

$$\ln y=\ln x^2-\ln(x+3)-\ln(2x+1)$$

根据(10.20),左侧对 $x$ 的导数为

$$\frac{d}{dx}(\text{左边})=\frac{1}{y}\frac{dy}{dx}$$

而右边的导数为

297

$$\frac{d}{dx}(\text{右边})=\frac{2x}{x^2}-\frac{1}{x+3}-\frac{2}{2x+1}=\frac{7x+6}{x(x+3)(2x+1)}$$

当两个结果相等,并均被 $y$ 相乘时,我们得到所求的导数为

$$\frac{dy}{dx}=\frac{7x+6}{x(x+3)(2x+1)}y$$

$$= \frac{7x+6}{x(x+3)(2x+1)} \frac{x^2}{(x+3)2x+1)}$$

$$= \frac{x(7x+6)}{(x+3)^2(2x+1)^2}$$

**例 10** 求 $y = x^a e^{kx-c}$ 的导数 $dy/dx$。对方程两边取自然对数,我们有

$$\ln y = a \ln x + \ln e^{kx-c} = a \ln x + kx - c$$

将两边对 $x$ 微分,并利用(10.20),则我们得到

$$\frac{1}{y}\frac{dy}{dx} = \frac{a}{x} + k$$

和
$$\frac{dy}{dx} = \left(\frac{a}{x} + k\right)y = \left(\frac{a}{x} + k\right)x^a e^{kx-c}$$

## 练习 10.5

---

1 求下列函数的导数:

(a) $y = e^{2t+4}$

(b) $y = e^{1-7t}$

(c) $y = e^{t^2+1}$

(d) $y = 3e^{2-t^2}$

(e) $y = e^{ax^2+bx+c}$

(f) $y = xe^x$

(g) $y = x^2e^{2x}$

(h) $h = axe^{bx+c}$

2 (a) 运用方程 $\ln at = \ln a + \ln t$ 验证例 3 的导数。

(b) 运用方程 $\ln t^c = c \ln t$ 验证例 4 的结果。

3 求下列函数的导数

(a) $y = \ln 8t^5$

(b) $y = \ln at^c$

(c) $y = \ln(t+9)$

(d) $y = 5\ln(t+1)^2$

(e) $y = \ln x - \ln(1+x)$

(f) $y = \ln[x(1-x)^8]$

(g) $y = \ln\left(\frac{3x}{1+x}\right)$

(h) $y = 5x^4 \ln x^2$

4　求下列函数的导数

$(a)\ y=5^t$　　$(d)\ y=\log_7 7x^2$

$(b)\ y=\log_2(t+1)$　　$(e)\ y=\log_2(8x^2+3)$

$(c)\ y=13^{2t+3}$　　$(f)\ y=x^2\log_3 x$

5　证明(10.21)的两个公式。

298

6　证明函数 $v=Ae^{rt}(A>0, r>0)$ 和函数 $A=ve^{-rt}(v>0,\ r>0)$ 均为单调函数，但一个单调递增，另一个单调递减，且二者均为严格凸函数(参见练习 10.2－5)。

7　首先对下列函数取自然对数后，再求其导数：

$(a)\ y=\frac{3x}{(x+2)(x+4)}$　　$(b)\ y=(x^2+3)e^{x^2+1}$

## 10.6　最优时间安排

我们所拥有的关于指数函数和对数函数的知识现在可以应用于某些简单的最优时间安排问题。

### 酒的窖藏问题

假设某酒商拥有特定数量的(比如一箱)的葡萄酒，他可以现在销售以得到 $K$ 美元，也可以将其窖藏某一时间以便以更高的价格销售。已知酒的增长值($V$)是时间的如下函数：

$$V=Ke^{\sqrt{t}}\quad [=K\exp(t^{1/2})] \tag{10.22}$$

从而若 $t=0$(现在销售)，则 $V=K$。假设窖藏成本为零，酒商的问题便是确定何时销售以得到最大利润。①

因酒的成本为酒商早已支付的“沉没”成本，且因假定不存在

① 考虑窖藏成本会使我们遇到我们现在还无法解决的困难。在 13 章中，我们会再研究这一问题。

窖藏成本,最大化利润亦即最大化销售收益,或者使 $V$ 值最大化。但有一点要注意,对应于每一特定时点 $t$ 的每一 $V$ 值表示在不同日期收到的美元数额,由于涉及利息因素,一个日期的 $V$ 值与另一日期的 $V$ 值不能直接比较。解决这一困难的办法是将每一 $V$ 值**贴现**到等价的**现值**(在时间 $t=0$ 的价值),则所有的 $V$ 值都具有了可比较的基础。

我们假设在连续复利基础上的利率水平是 $r$。则根据(10.11),$V$ 的现值可以表示为

$$A(t) = Ve^{-rt} = Ke^{\sqrt{t}}e^{-rt} = Ke^{\sqrt{t}-rt} \qquad (10.22')$$

其中 $A$ 表示现值 $V$,它是 $t$ 的函数。因此,我们的问题相当于求使 $A$ 最大化的 $t$ 值。

## 299 最大化条件

最大化 $A$ 的一阶条件是令 $dA/dt=0$。为求此导数,我们可以将(10.22′)直接对 $t$ 微分,或者间接地,先对(10.22′)两边取自然对数,然后再将其对 $t$ 求微分。我们先描述后一种方法。

首先,由(10.22′)我们得到方程

$$\ln A(t) = \ln K + \ln e^{\sqrt{t}-rt} = \ln K + (t^{1/2} - rt)$$

将两边对 $t$ 求导,则我们有

$$\frac{1}{A}\frac{dA}{dt} = \frac{1}{2}t^{-1/2} - r$$

或

$$\frac{dA}{dt} = A\left(\frac{1}{2}t^{-1/2} - r\right)$$

因为 $A \neq 0$,当且仅当

$$\frac{1}{2}t^{-1/2} = r \quad 或 \quad \frac{1}{2\sqrt{t}} = r \quad 或 \quad \frac{1}{2r} = \sqrt{t}$$

条件 $dA/dt=0$ 方能得到满足。这意味着最优窖藏时间长度为

$$\bar{t}=\left(\frac{1}{2r}\right)^2=\frac{1}{4r^2}$$

例如,若 $r=0.10$,则 $\bar{t}=25$,即此酒商应将这箱酒窖藏 25 年。注意,利息率(贴现率)越高最窖藏期越短。

一阶条件 $1/(2\sqrt{t})=r$,具有一个简单的经济解释。左边的表达式仅表示酒的价值 $V$ 的增长率,因为由(10.22)

$$\frac{dV}{dt}=\frac{d}{dt}K\exp(t^{1/2})=K\frac{d}{dt}\exp(t^{1/2})\quad [K\text{ 常数}]$$

$$=K\left(\frac{1}{2}t^{-1/2}\right)\exp(t^{1/2})\qquad [\text{由}(10.20)]$$

$$=\left(\frac{1}{2}t^{-1/2}\right)V\qquad [\text{由}(10.22)]$$

所以 $V$ 的增长率实际上是一阶条件中的左边的表达式:

$$r_V\equiv\frac{dV/dt}{V}=\frac{1}{2}t^{-1/2}=\frac{1}{2\sqrt{t}}$$

与其相对照,右边表达式 $r$,则是利息率或若酒现在卖掉,可收到现金的复利增长率(相对于窖藏的机会成本)。因此,如图 10.4 所示,令两个瞬时增长率相等,也就意味着保存酒直至窖藏的好处完全消失,亦即保存酒直至当酒的价值的递减的增长率与销售酒所得现金的不变的利息率相等时止。 300

下一步要做的工作是检验 $\bar{t}$ 值是否满足 $A$ 最大化的二阶条件。$A$ 的二阶导数为

$$\frac{d^2A}{dt^2}=\frac{d}{dt}A\left(\frac{1}{2}t^{-1/2}-r\right)=A\frac{d}{dt}\left(\frac{1}{2}t^{-1/2}-r\right)$$

$$+\left(\frac{1}{2}t^{-1/2}-r\right)\frac{dA}{dt}$$

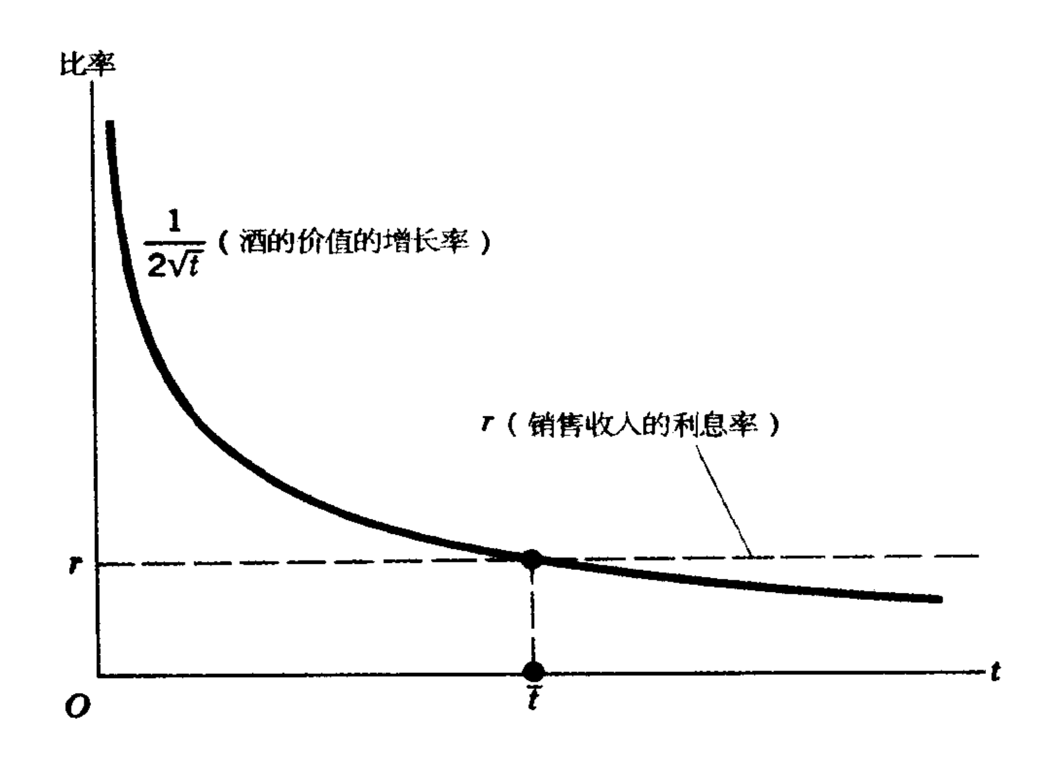



图 10.4

但因当我们在均衡(最优)点计值时，$dA/dt=0$，最后一项可舍弃，所以我们有

$$\frac{d^2A}{dt^2}=A\frac{d}{dt}\left(\frac{1}{2}t^{-1/2}-r\right)=A\left(-\frac{1}{4}t^{-3/2}\right)=\frac{-A}{4\sqrt{t^3}}$$

考虑到 $A>0$，当在 $\bar{t}>0$ 计值时二阶导数为负，因而保证了解值 $\bar{t}$ 确实使利润最大化。

## 伐木问题

另一个问题是伐木问题。它也涉及到采取行动的最优时间选择。

假定木材(早已在给定土地上种植)的价值时间的如下增函数：

$$V=2^{\sqrt{t}}$$

单位为千美元。假设贴现率为 $r$（在连续的基础上），并假定在树木成长期间养护成本为零，什么是砍伐树木并进行销售的最佳时机？

像在葡萄酒中的问题一样，我们首先将 $V$ 变换成现值： 301

$$A(t)=Ve^{-rt}=2^{\sqrt{t}}e^{-rt}$$

因此 $\ln A=\ln 2^{\sqrt{t}}+\ln e^{-rt}=\sqrt{t}\ln 2-rt=t^{1/2}\ln 2-rt$

要使 $A$ 最大化，我们必须令 $dA/dt=0$。将 $\ln A$ 对 $t$ 求导，并以 $A$ 与之相乘，可以得到一阶导数：

$$\frac{1}{A}\frac{dA}{dt}=\frac{1}{2}t^{-1/2}\ln 2-r$$

因此
$$\frac{dA}{dt}=A\left(\frac{\ln 2}{2\sqrt{t}}-r\right)$$

因 $A\neq 0$，当且仅当

$$\frac{\ln 2}{2\sqrt{t}}=r \quad 或 \quad \sqrt{t}=\frac{\ln 2}{2r}$$

时，条件 $dA/dt=0$ 方能得到满足。从而，最优生长年数为

$$\bar{t}=\left(\frac{\ln 2}{2r}\right)^2$$

很明显，贴现率越高，树木越应早砍伐。

为保证 $\bar{t}$ 是最大化解而非最小化解，应进行二阶条件检验。但这个问题留给读者作为练习之用。

在本例中，我们通过假设树木为以前所植，从而舍去了植树成本，使沉没的植树成本合理地排除在最优决策问题之外。但若决策不是关于何时收获的问题，而是关于是否植树的问题，则现在发生的植树成本就必须与树木产出的现值相比较，且现值应按最优值 $\bar{t}$ 来计算。例如，若 $r=0.05$，则有

$$\bar{t} = \left(\frac{0.6931}{0.10}\right)^2 = (6.931)^2 = 48.0 \text{ 年}$$

和
$$\bar{A} = 2^{6.931} e^{-0.05(48.0)} = (122.0222) e^{-2.40}$$
$$= 122.0222(0.0907) = \$ 11.0674(\text{千})$$

所以仅当植树成本低于 $\bar{A}$ 时，植树才是值得的（同样假定养护成本为零）。

## 练习 10.6

---

1 若酒的价值按函数 $V = Ke^{2\sqrt{t}}$ 增长，而不是按（10.22）式增长，那么，302 酒商储酒时间应为多长？

2 检验伐木问题的二阶条件。

3 作为本节描述的最优化问题的推广，证明：

（$a$）有任意值函数 $V = f(t)$，已知连续贴现率 $r$，$V$ 的现值 $A$ 达到极大值的一阶条件是 $V$ 的增长率等于 $r$。

（$b$）最大值的二阶充分条件实际上相当于规定 $V$ 的增长率随时间递减。

---

## 10.7 指数函数与对数函数导数的进一步应用

除了在最优化问题中应用以外，10.5 节的导数公式在经济中还有其它有意义的应用。

## 求增长率

当变量 $y$ 是时间的函数,即 $y=f(t)$,其瞬时增长率定义为[①]

$$r_y \equiv \frac{dy/dt}{y} = \frac{f'(t)}{f(t)} = \frac{\text{边际函数}}{\text{总函数}} \tag{10.23}$$

但由(10.20)可知,此比率恰好是 $\ln f(t)=\ln y$ 的导数。因此,为求时间函数 $f(t)$ 瞬时增长率,我们可以不采取将函数对 $t$ 求导,然后再以 $f(t)$ 相除的方法,而是简单地将函数两边取自然对数,然后再将 $\ln f(t)$ 对时间求导。[②] 后一种方法可能是一种更简单的方法,若 $f(t)$ 是积或商的表达式,通过取对数,会使其简化为可加项的和或差。

**例 1** 求 $V=Ae^{rt}$ 的增长率,其中 $t$ 表示时间。我们已知道 $V$ 的增长率即 $r$,但我们还是通过求 $\ln V$ 的导数来对此进行检验:

$$\ln V = \ln A + rt \ln e = \ln A + rt \qquad [A \text{ 常数}]$$

因此

$$r_V = \frac{d}{dt}\ln V = 0 + \frac{d}{dt}rt = r$$

结论得证。

**例 2** 求 $y=4^t$ 的增长率。在本例中

303

$$\ln y = \ln 4^t = t \ln 4$$

因此

$$r_y = \frac{d}{dt}\ln y = \ln 4$$

---

① 若变量 $t$ 不表示时间,那么,表达式 $(dy/dt)/y$ 被称作 $y$ 对 $t$ 变化的比率。

② 若我们在一个两维空间中绘出函数 $f(t)$ 对 $t$ 的图形,曲线的斜率便会告知我们 $f(t)$ 的增长率。这还证明了所谓"半对数刻度"图的合理性。

它确实应当如此，因为 $e^{\ln 4} \equiv 4$，因而，$y=4^t$ 可改写成 $y=e^{(\ln 4)t}$，由此我们马上可看出(ln4)是 $y$ 的增长率。

## 组合函数的增长率

为此讨论深化一步，我们考察两个时间函数之积的增长率：

$$y = uv \qquad \text{其中} \begin{cases} u = f(t) \\ v = g(t) \end{cases}$$

取 $y$ 的自然对数，我们得到

$$\ln y = \ln u + \ln v$$

因此所求增长率为

$$r_y = \frac{d}{dt}\ln y = \frac{d}{dt}\ln u + \frac{d}{dt}\ln v$$

但等号右边两项分别为 $u$ 和 $v$ 的增长率。因此我们有法则

$$r_{(uv)} = r_u + r_v \tag{10.24}$$

以文字表达即：两函数积的瞬时增长率等于每个函数增长率的和。

通过类似的步骤还可以证明，函数商的增长率等于被除数与除数两者增长率之差(参见练习 10.7－4)：

$$r(u/v) = r_u - r_v \tag{10.25}$$

**例3** 若消费 $C$ 以比率 $\alpha$ 增长，人口 $H$ 以比率 $\beta$ 增长，那么人均消费增长率为多少？因人均消费等于 $C/H$，其增长率应为

$$r_{(C/H)} = r_C - r_H = \alpha - \beta$$

现在我们来考察两个时间函数的和的瞬时增长率：

$$z = u + v \qquad \text{其中} \begin{cases} u = f(t) \\ v = g(t) \end{cases}$$

304 这时，自然对数为

$$\ln z = \ln(u+v) \quad [\neq \ln u + \ln v]$$

因此,

$$r_z = \frac{d}{dt}\ln z = \frac{d}{dt}\ln(u+v)$$

$$= \frac{1}{u+v}\frac{d}{dt}(u+v) \qquad [由(10.20)]$$

$$= \frac{1}{u+v}[f'(t)+g'(t)]$$

但由(10.23),我们有 $r_u = f'(t)/f(t)$,从而 $f'(t) = f(t)r_u = ur_u$。类似地,我们有 $g'(t) = v\, r_v$。因此,我们可以写出法则

$$r_{(u+v)} = \frac{u}{u+v}r_u + \frac{v}{u+v}r_v \tag{10.26}$$

此式表明,两个时间函数和的增长率是每个函数增长率的加权平均值。

同理,我们有(见练习 10.7-5)

$$r_{(u-v)} = \frac{u}{u-v}r_u - \frac{v}{u-v}r_v \tag{10.27}$$

**例 4** 某国商品出口为 $G = G(t)$,其增长率为 $a/t$;该国服务出口为 $S = S(t)$,其增长率为 $b/t$,总出口增长率为多少? 因为总出口 $X_{(t)} = G_{(t)} + S_{(t)}$,所以其增长率应为

$$r_X = \frac{G}{X}r_G + \frac{S}{X}r_S$$

$$= \frac{G}{X}\left(\frac{a}{t}\right) + \frac{S}{X}\left(\frac{b}{t}\right) = \frac{Ga+Sb}{Xt}$$

## 求点弹性

我们已知道,给定 $y = f(t)$,$\ln y$ 的导数度量 $y$ 的瞬时增长率。

现在我们来考察给定函数 $y=f(x)$，当我们将($\ln y$)对($\ln x$)求导，而不是对 $x$ 求导时，会出现何种情况。

首先，我们定义 $u \equiv \ln y$ 和 $v \equiv \ln x$。则我们可以看到将 $u$ 与 $y$ 以及 $x$ 与 $v$ 联系起来的链式关系：

$$u \equiv \ln y \qquad y=f(x) \qquad x \equiv e^{\ln x} \equiv e^{v}$$

因而，($\ln y$)对($\ln x$)的导数为

$$\frac{d(\ln y)}{d(\ln x)}=\frac{du}{dv}=\frac{du}{dy}\frac{dy}{dx}\frac{dx}{dv}$$

$$=(\frac{d}{dy}\ln y)(\frac{dy}{dx})(\frac{d}{dv}e^{v})=\frac{1}{y}\frac{dy}{dx}e^{v}=\frac{1}{y}\frac{dy}{dx}x=\frac{dy}{dx}\frac{x}{y}$$

305 此式即函数的点弹性。因此我们确立一个一般原则：对于函数 $y=f(x)$，$y$ 对 $x$ 的点弹性为

$$\varepsilon_{yx}=\frac{d(\ln y)}{d(\ln x)} \tag{10.28}$$

应注意，符号中的下标 $yx$ 仅表明 $y$ 与 $x$ 是所涉及的两个变量，而不意味着 $y$ 与 $x$ 的乘积。这与 $r_{(uv)}$ 中的情形是不一样的，($uv$)确实表示乘积关系。同样，我们现在还可以用另一种方法求函数的点弹性，即用对数法，若给定函数以积或商的形式出现，对数法通常是一种更为简单的方法。

**例 5** 给定 $Q=k/P$，其中 $k$ 为正常数，求需求的点弹性。这是一个等轴双曲线方程(见图 2.8$d$)，而且，众所周知，此类方程在所有点具有单位点弹性，为证明这一点，我们应用(10.28)。因需求函数的自然对数为

$$\ln Q=\ln k-\ln P$$

所以($Q$ 对 $P$ 的)需求点弹性为

$$\varepsilon_d = \frac{d(\ln Q)}{d(\ln P)} = -1 \qquad 或 \qquad |\varepsilon_d| = 1$$

(10.28)中的结果是运用链式求导法则推导出来的。有趣的是,类似的链式法则对弹性也成立,即给定函数 $y = g(w)$,其中 $w = h(x)$,我们有

$$\varepsilon_{yx} = \varepsilon_{yw}\varepsilon_{wx} \tag{10.29}$$

证明如下:

$$\varepsilon_{yw}\varepsilon_{wx} = \left(\frac{dy}{dw}\frac{w}{y}\right)\left(\frac{dw}{dx}\frac{x}{w}\right) = \frac{dy}{dw}\frac{dw}{dx}\frac{w}{y}\frac{x}{w} = \frac{dy}{dx}\frac{x}{y} = \varepsilon_{yx}$$

## 练习 10.7

---

1 求瞬时增长率

(a) $y = 3t^2$ (c) $y = ab^t$ (e) $y = t/3^t$

(b) $y = at^c$ (d) $y = 2^t(t^2)$

2 若人口增长函数为 $H = H_0(2)^{bt}$,消费函数为 $C = C_0e^{at}$,运用自然对数法求人口增长率、消费增长率、人均消费增长率。

3 若 $y$ 通过 $y = x^k$ 与 $x$ 相联系,那么,增长率 $r_y$ 与 $r_x$ 以何种方式相联系?



4 证明:若 $y = u/v$,其中 $u = f(t)$, $v = g(t)$,则如(10.25)所示,$y$ 的增长率将为 $r_y = r_u - r_v$。

5 证明增长率法则(10.27)。

6 给定需求函数 $Q_d = k/P^n$,其中 $k$ 与 $n$ 均为正常数,运用(10.28)求需求点弹性$\varepsilon_d$(见练习 8.1-5)。

7 (a)已知 $y = wz$,其中 $w = g(x)$, $z = h(x)$,证明:

$\varepsilon_{yx} = \varepsilon_{wx}\varepsilon_{zx}$

(b)已知 $y = u/v$,其中 $u = G(x)$, $v = H(x)$,证明:

$\varepsilon_{yx} = \varepsilon_{ux}\varepsilon_{vx}$

8　已知 $y=f(x)$，证明导数

$$\frac{d(\log_b y)}{d(\log_b x)} \qquad \text{（对数的底为 } b \text{ 而非 } e\text{）}$$

也度量点弹性 $\varepsilon_{yx}$。

9　若货币需求 $M_d$ 是国民收入 $Y=Y(t)$ 的函数，利息率为 $i=i(t)$，证明：$M_d$ 的增长率可以表示成 $r_Y$ 与 $r_i$ 的加权和，

其中，权数分别为 $M_d$ 对 $Y$ 与 $i$ 的弹性。

$$r_{M_d}=\varepsilon_{M_d Y}r_Y+\varepsilon_{M_d i}r_i$$

10　已知生产函数 $Q=F(K,L)$，求用 $K$ 和 $L$ 的增长率表示的 $Q$ 的增长率的一般表达式。

---

# 第11章 多于一个选择变量的情况

在第9章中讨论的最优化问题一直限定在目标函数具有单一选择变量的框架内。在上一章,讨论的范围扩展到了指数目标函数,但所涉及的仍是单一选择变量问题。现在我们必须研究一种求包括两个或两个以上选择变量目标函数的相对极值的方法。只有这样,我们才能解决这些即将面临的问题,比如多产品厂商利润最大化的决策问题。这些决策包括几种商品最优产出水平及几种不同投入最优组合的选择等问题。

307

为利用目标函数的图形,我们首先讨论含两个选择变量的目标函数 $z=f(x,y)$ 的情况。然后再将分析结果推广至难以作图的 $n$ 个变量的情况。但是,无论变量数量为多少,我们一般假定,当目标函数以一般形式写出时,它总具有连续的至任意所需阶数的偏导数。这一假定将保证目标函数及其偏导数的平滑性及可微性。

对于多变量的函数,其极值也分绝对或整体极值,以及相对或局部极值两类。同以前一样,我们的注意力主要集中于相对极值,为此,我们将省略形容词“相对”二字;一般所说的极值,均指相对极值,除非另作说明。但在11.5节中,我们将给绝对极值的条件以适当的考虑。

## 11.1 最优化条件的微分形式

308 第9章对单一选择变量函数最优化条件问题的讨论完全是以导数概念来表达的。为讨论两个或多个变量问题作准备,知道如何按照微分概念来表达最优化条件也是有益的。

### 一阶条件

考察函数 $z=f(x)$,其图形如图11.1所示。在极大值点 $A$ 和极小值点 $B$,$z$ 值必然是稳定的。换句话说,当 $x$ 变化时,$dz$ 瞬间等于零,是 $z$ 存在相对极值的必要条件。这一条件便构成了相对极值一阶条件的微分形式。尽管 $dz=0$ 是必要条件,但它显然既不是极大值,也不是极小值的充分条件,因为在图11.1中的拐点 $C$ 也具有 $dz=0$ 的性质。

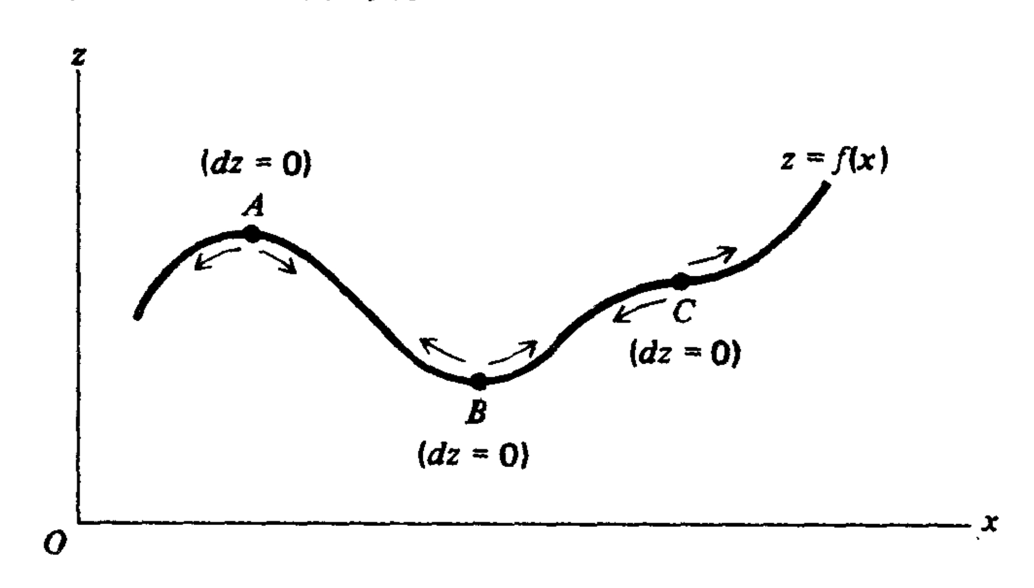



图 11.1

为证明上述条件与一阶条件的导数形式 $dz/dx=0$ 或 $f'(x)=0$ 是等价的,我们回顾一下 $z=f(x)$ 的微分为

$$dz = f'(x)dx \tag{11.1}$$

我们注意到，当 $x$ 无变化，即 $dx=0$ 时，$dz$ 自动为零。当然这并非一阶条件的全部含义。一阶条件所要求的是当 $x$ 变化时 $dz$ 为零，即 $x$ 发生任意（正、负，但不为零）无穷小变化时，$dz$ 为零。在 $dx\neq 0$ 的情况下，当且仅当 $f'(x)=0$ 时，$dz$ 能够为零。因此，导数条件 $f'(x)=0$ 与微分条件"对任意 $dx$ 不为零，$dz=0$"，实际上是等价的。

## 二阶条件

309

极值点，比如图 11.1 中的 $A$ 点，在图形上具有这样的性质：当我们沿曲线以无穷小变化向 $A$ 点的左侧（$dx<0$）和右侧（$dx>0$）滑动时，$z$ 在两个方向上均是递减的。达到这一点的充分条件是在 $A$ 点的最小邻域内，在 $A$ 点左右两边均有 $dz<0$。[①] $dz$ 在 $A$ 点等于零，在 $A$ 点两侧小于零这一事实，意味着无论从哪个方向移离 $A$ 点，$dz$ 必然都是递减的。换言之，此条件相当于对任意 $dx$ 不为零，$d(dz)<0$ 或用更简单的符号表示：$d^2z<0$。符号 $d^2z=d(dz)$ 表示微分的微分，称之为 $z$ 的二阶微分。上述关于 $d^2z$ 的条件构成了极大值的二阶充分条件的微分形式。

注意，$d^2z$ 为负是极大值的充分条件，但不是必要条件。理由是，在某些情况下，在 $z$ 的极大值，$d^2z$ 可能恰好为零而不为负。当然这种可能性极易使我们联想起 $N$ 阶导数检验的情况，其中极

---

① 借助于(11.1)，这一点便可以明确。令 $A$ 点两侧 $dz<0$，则 $f(x)$ 与 $dx$ 符号必然相反。这意味着在 $A$ 点的左侧（令 $dx<0$），$f'(x)$ 必定为正，所以 $f$ 曲线必然向上方倾斜。类似地，在 $A$ 点的右侧（令 $dx>0$），$f'(x)$ 必定为负，所以 $f$ 曲线必定向下方倾斜。因此，点 $A$ 是峰顶。

大值可能具有二阶导数为零的特征。实际上，在单变量函数情况下，二阶微分 $d^2z$ 的符号与二阶导数 $d^2z/dx^2$ 或者 $f''(x)$ 的符号之间存在着非常密切的关系。我们下面就将对这种关系进行阐述。

已知 $dz=f'(x)dx$，仅通过 $dz$ 的进一步微分，便可以得到 $d^2z$。但在微分时我们必须记住，$dx$，在这里代表 $x$ 的任意的或给定的非零变化，在微分过程中应当作常数看待。因而，$dz$ 仅随 $f'(x)$ 的变化而变化，但因 $f'(x)$ 又是 $x$ 的函数，所以在最终分析中，$dz$ 仅随 $x$ 变化而变化。考虑到这一点，我们有

$$\begin{aligned} d^2z \equiv d(dz) &= d[f'(x)dx] && [\text{由}(11.1)] \\ &= [df'(x)]dx && [dx\ \text{是常数}] \qquad (11.2) \\ &= [f''(x)dx]dx = f''(x)dx^2 \end{aligned}$$

注意，在(11.2)中，指数 2 以两种完全不同的方式出现。在符号 $d^2z$ 中，指数 2 表示 $z$ 的二阶微分；而在符号 $dx^2 \equiv (dx)^2$ 中，指数2 表示一阶微分 $dx$ 的平方。(11.2)的结果给出了 $d^2z$ 与 $f''(x)$ 之间的直接联系。由于我们现在仅考察 $dx$ 的非零值，所以 $dx^2$ 项总为正，因而 $d^2z$ 和 $f''(x)$ 必然取相同的代数符号。

这一事实进一步支持了我们前面的这一观点：作为 $z$ 存在极值的充分条件，“对于任意非零 $dx$ 值，微分条件 $d^2z<0$”与导数条件 $f''(x)<0$ 是等价的。转向 $z$ 的极小值的情况，由(11.2)我们还可以看到，充分导数条件 $f''(x)>0$ 与“对于任意非零的 $dx$，$d^2z>0$”的微分条件是等价的。最后，我们可以由(11.2)推断 $z$ 的极值的二阶必要条件：

310

对于 $z$ 的极大值：$f''(x)\leq 0$

对于 $z$ 的极小值：$f''(x)\geq 0$

可以分别变换成：

$$\left.\begin{array}{l}\text{对于 } z \text{ 的极大值}: d^2z \leqslant 0 \\ \text{对于 } z \text{ 的极小值}: d^2z \geqslant 0\end{array}\right\} \text{对于 } dx \text{ 的任意非零值。}$$

**微分条件与导数条件**

现在我们已证明了以 $dz$ 和 $d^2z$ 来表示导数形式的一阶条件和二阶条件的可行性。读者可能要问：在导数条件已经适用的情况下，为什么我们还要推出一套微分条件呢？答案是，微分条件是以这样一种形式表述的：它可以直接从单变量的情况推广至两个或两个以上变量的情况，而导数形式则不能作这种推广。更具体地说，只要短语"对于 $dx$ 的任意非零值"作适当修改以反映选择变量数量的变化，一阶条件（$dz$ 为零）和二阶条件（$d^2z$ 为正或负）可以有效地应用于所有情况。

但是，这并不意味着导数条件不能发挥进一步的作用了。相反，因为导数条件更便于应用，在运用微分条件完成向更多变量推广的过程之后，我们将努力发展并利用适于那些情况的导数条件。

## 11.2 两个变量函数的极值

对于一个选择变量的函数，在二维空间中，极值在图形上是以峰顶或谷底来表示的。具有两个选择变量的函数 $z = f(x, y)$ 的图形，是三维空间中的一个曲面；其极值仍与峰顶或谷底相联系，但其"峰顶"或"谷底"现已具有三维空间的特征。在这个新的内容中，其形状分别像圆丘和碗。在图 11.2 中的两个图形对此进行了描述。图 $a$ 中的点 A，圆丘之顶，构成了极大值；此点的 $z$ 值大于

其紧邻邻域中的其它任意值。类似地,图 $b$ 中的点 $B$,即“碗底”表示极小值;在其紧邻邻域中的任意函数值均大于该点的函数值。

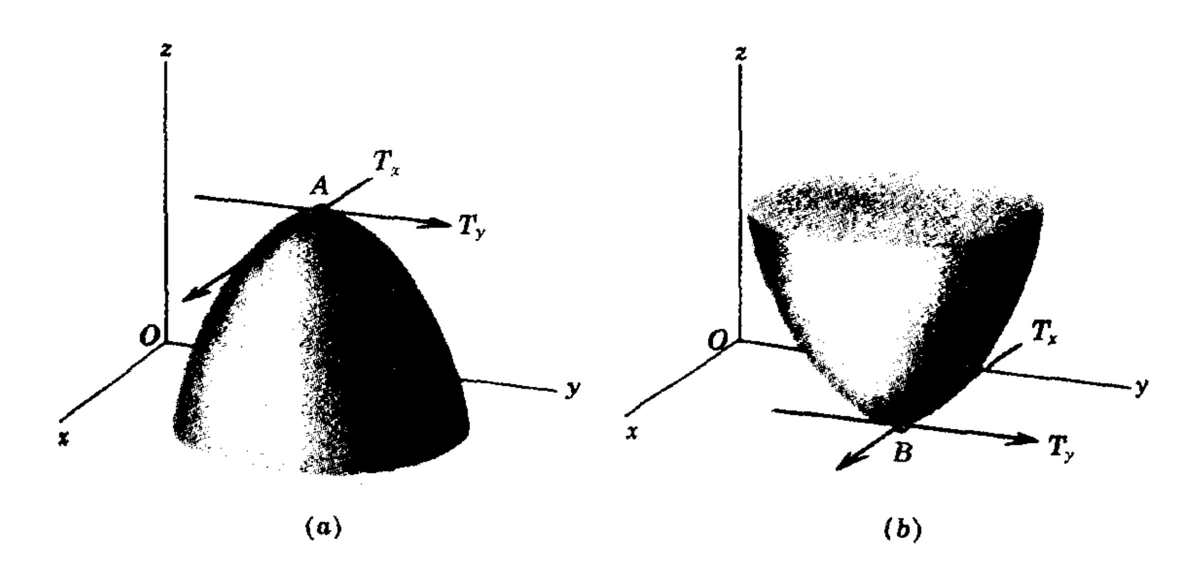



图 11.2

## 311 一阶条件

对于函数

$$z=f(x,y)$$

极值(极大值或极小值)的一阶必要条件仍是 $dz=0$。但因这里有两个自变量,所以现在 $dz$ 是全微分,因此一阶条件应修正为如下形式:

对于二者不同时为零的任意 $dx$ 和 $dy$ 值,

$$dz=0 \tag{11.3}$$

关于(11.3)合理性的解释与单变量情况下对条件 $dz=0$ 的解释是类似的:极值点必定为稳定点,而在稳定点,对两个变量 $x$ 与 $y$ 的任意无穷小变化,$z$ 必定为常数。

在目前两变量情况下,全微分为

$$dz=f_x dx+f_y dy \tag{11.4}$$

要满足条件(11.3),两个偏导数 $f_x$ 和 $f_y$ 同时为零是充要条件。因此,一阶条件(11.3)的等价的导数形式为

$$f_x = f_y = 0 \quad \left[\text{或}\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} = 0\right] \tag{11.5}$$

此条件存在一个简单的图形解释。参看图 11.2 中的点 $A$,使 $f_x$ 在 $A$ 点为零,意味着通过 $A$ 点且与 $xz$ 平面平行的切线 $T_x$ 的斜率必然为零。同理,使 $f_y$ 在 $A$ 点等于零,意味着通过 $A$ 点且与 $yz$ 平面($x$ 保持不变)平行的切线 $T_y$ 的斜率必然为零。读者还可以

312

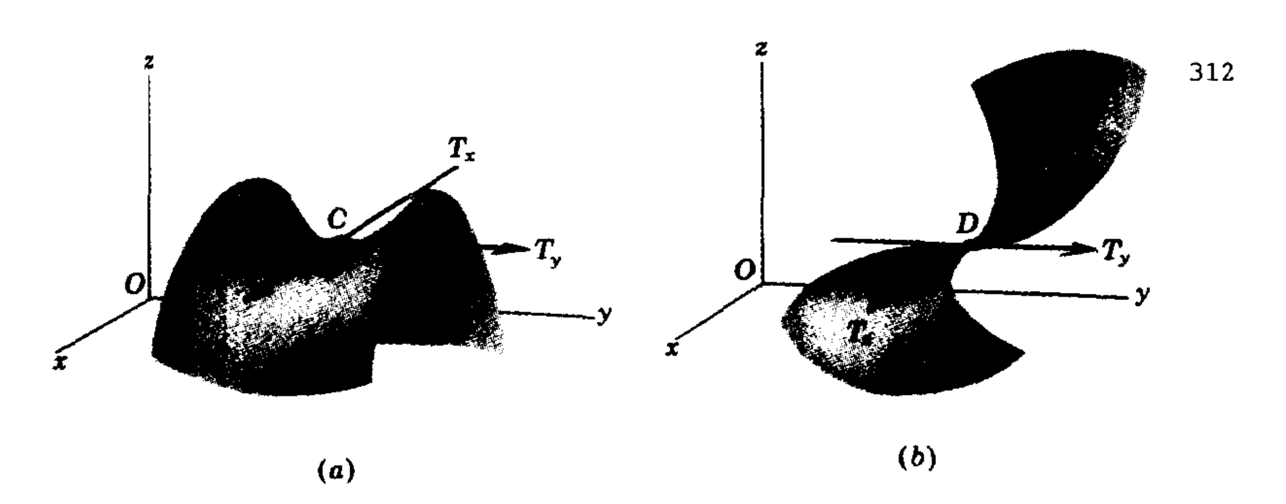



图 11.3

验证,这些切线要求也可以应用于图 11.2$b$ 中的极小值点 $B$。这是因为条件(11.5)同(11.3)一样,也是极大值和极小值存在的必要条件。

同以往的讨论一样,一阶条件是极值存在的必要条件,但不是充分条件。在图 11.3 的两个图形中也可以看到,仅靠一阶条件不足以确定一个极值。在图 $a$ 中的 $C$ 点,$T_x$ 和 $T_y$ 斜率均为零,但此点并不是合格的极值点:以 $yz$ 平面为背景来考察,它是一个极小值;而从 $xz$ 平面来考察,它又是一个极大值。由于图形上的原因,这种具有"双重人格"的点,被称作鞍点。类似地,图 11.3$b$ 中

的点 $D$,尽管具有水平线 $T_x$ 和 $T_y$ 的特征,也不是极值点。它位于扭曲的曲面上,无论从 $xy$ 平面,还是从 $yz$ 平面上看,它都是一个拐点。这些反例从根本上排除了一阶条件作为极值充分条件的可能性。

为导出极值存在的充分条件,我们需要考察与二阶偏微分相联系的二阶全导数。

## 二阶偏导数

函数 $z=f(x,y)$ 可以引致两个一阶偏导数

$$f_x \equiv \frac{\partial z}{\partial x} \quad 和 \quad f_y \equiv \frac{\partial z}{\partial y}$$

因 $f_x$ 本身是 $x$(及 $y$)的函数,所以当 $y$ 保持不变时,我们可以用特定的二阶偏导数 $f_{xx}$ 或 $\partial^2 z/\partial x^2$ 来表示 $f_x$ 对 $x$ 的变化率:

$$f_{xx} \equiv \frac{\partial}{\partial x}(f_x) \quad 或 \quad \frac{\partial^2 z}{\partial x^2} \equiv \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right)$$

313 符号 $f_{xx}$ 具有双下标,表示原函数 $f$ 对 $x$ 求偏微分两次,而除了使用偏导数符号外,符号 $\partial^2 z/\partial x^2$ 与 $d^2 z/dx^2$ 是类似的。我们可以以完全类似的方式运用二阶偏导数

$$f_{yy} \equiv \frac{\partial}{\partial y}(f_y) \quad 或 \quad \frac{\partial^2 z}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right)$$

来表示当 $x$ 固定不变时,$f_y$ 对 $y$ 的变化率。

但是,我们知道 $f_x$ 也是 $y$ 的函数,$f_y$ 也是 $x$ 的函数,因此,可以写出另外两个偏导数:

$$f_{xy} \equiv \frac{\partial^2 z}{\partial x \partial y} \equiv \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right) \quad 和 \quad f_{yx} \equiv \frac{\partial^2 z}{\partial y \partial x} \equiv \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right)$$

它们被称作交叉(或混合)偏导数,因为它们度量一个一阶偏导数

对“另一个”变量的变化率。

值得重复一下：$z=f(x,y)$的二阶偏导数，像$z$、$f_x$及$f_y$一样，都是变量$x$和$y$的函数。当需要强调这一事实时，我们可以将$f_{xx}$写成$f_{xx}(x,y)$，将$f_{xy}$写成$f_{xy}(x,y)$等。同样，我们可以使用符号$f_{yx}(1,2)$来表示$f_{yx}$在$x=1,y=2$计算的值，等等。

尽管$f_{xy}$与$f_{yx}$是分别定义的，但按照通称为杨氏定理（*Young's theorem*）的命题，二者是相等的，只要两个交叉偏导数是连续的。在此情况下，由于$f_{xy}=f_{yx}$，所以求偏微分的顺序变量不重要了。对于我们所研究的一般类型的具体函数，连续性条件通常是满足的；对于我们前面所提到的那些一般函数，我们总是假定连续性条件成立。所以一般情况下我们可以预期能够求得相同的交叉偏导数。事实上，此定理还可以应用于三个或更多变量的函数。例如，已知$z=g(u,v,w)$，假设这些偏导数全都连续，则其混合偏导数具有$g_{uv}=g_{vu}$，$g_{vw}=g_{wv}$等特征。

**例1** 求函数

$$z=x^3+5xy-y^2$$

四个二阶偏导数。此函数的一阶偏导数为：

$$f_x=3x^2+5y \quad 和 \quad f_y=5x-2y$$

因此，通过进一步求导，我们有

$$f_{xx}=6x，\ f_{yx}=5，\ f_{xy}=5，\ f_{yy}=-2$$

正如所预期的那样，$f_{yx}$与$f_{xy}$是相等的。

**例2** 求函数

$$z=x^2e^{-y}$$

的所有二阶偏导数。在本例中，一阶偏导数为：

$$f_x=2xe^{-y} \quad 和 \quad f_y=-x^2e^{-y}$$

314 因此，我们有

$$f_{xx}=2e^{-y} \quad f_{yx}=-2xe^{-y} \quad f_{xy}=-2xe^{-y} \quad f_{yy}=x^2e^{-y}$$

同样，我们看到 $f_{yx}=f_{xy}$。

注意，二阶偏导数都是原始变量 $x$ 和 $y$ 的函数。这一事实在例 2 中非常明显；即便在例 1 中也同样成立，尽管在该例中有些二阶偏导数恰好为常函数。

## 二阶全微分

给定(11.4)中的全微分 $dz$，根据我们掌握的二阶偏导数的概念，通过进一步微分 $dz$，我们可以推导出二阶全微分 $d^2z$ 的表达式。在推导时我们应记住，在方程 $dz=f_xdx+f_ydy$ 中，符号 $dx$ 和 $dy$ 表示 $x$ 与 $y$ 的任意或给定的变化；所以在微分过程中必须视为常数。因而 $dz$ 仅依赖于 $f_x$ 和 $f_y$，又因 $f_x$ 和 $f_y$ 本身又是 $x$ 和 $y$ 的函数，所以 $dz$ 像 $z$ 本身一样，也是 $x$ 和 $y$ 的函数。

为得到 $d^2z$，我们仅需应用如(11.4)所示的微分定义于 $dz$ 本身即可。因而，

$$\begin{aligned}
d^2z \equiv d(dz) &= \frac{\partial(dz)}{\partial x}dx+\frac{\partial(dz)}{\partial y}dy \quad [\text{参见}(11.4)] \\
&= \frac{\partial}{\partial x}(f_x dx+f_y dy)dx+\frac{\partial}{\partial y}(f_x dx+f_y dy)dy \\
&= (f_{xx}dx+f_{xy}dy)dx+(f_{yx}dx+f_{yy}dy)dy \\
&= f_{xx}dx^2+f_{xy}dydx+f_{yx}dxdy+f_{yy}dy^2 \\
&= f_{xx}dx^2+2f_{xy}dxdy+f_{yy}dy^2 \quad [f_{xy}=f_{yx}]
\end{aligned} \tag{11.6}$$

还要注意，在(11.6)中的指数 2 以两种不同方式出现。在符号

$d^2z$ 中，指数 2 表示 $z$ 的二阶全微分；而在符号 $dx^2 \equiv (dx)^2$ 中，指数表示一阶微分 $dx$ 的平方。

(11.6)中的结果表明按照给定的 $dx$ 和 $dy$ 表示 $d^2z$($dz$ 的变化)的大小，其中 $dx$ 和 $dy$ 是在定义域中某点($x_0, y_0$)处度量的。但为了计算 $d^2z$，也需知道二阶偏导数 $f_{xx}$、$f_{xy}$、$f_{yy}$ 的大小，它们均在点($x_0, y_0$)处计值——正如由(11.4)计算 $dz$，需要知道一阶偏导数的值一样。

**例 3** 给定 $z = x^3 + 5xy - y^2$，求 $dz$ 和 $d^2z$。此函数与例 1 相同。因此，将我们在例 1 中已得到的各导数代入(11.4)和 315 (11.6)，我们求得①

$$dz = (3x^2 + 5y)dx + (5x - 2y)dy$$

和

$$d^2z = 6xdx^2 + 10dxdy - 2dy^2$$

例如在点 $x = 1, y = 2$，我们有

$$dz = 13dx + dy \quad 和 \quad d^2z = 6dx^2 + 10dxdy - 2dy^2$$

并且对于从定义域中点 $x = 1, y = 2$ 处度量的给定的 $dx$ 与 $dy$，$dz$ 的符号说明 $z$ 的变化方向，而 $d^2z$ 的符号则揭示 $dz$ 是递增($d^2z > 0$)还是递减($d^2z < 0$)。

---

① 得到这一结果的另一种方式是直接对函数求微分：

$$\begin{aligned} dz &= d(x^3) + d(5xy) - d(y^2) \\ &= 3x^2dx + 5ydx + 5xdy - 2ydy \end{aligned}$$

记住 $dx$、$dy$ 为常数，对 $dz$ 再微分则有：

$$\begin{aligned} d^2z &= d(3x^2)dx + d(5y)dx + d(5x)dy - d(2y)dy \\ &= (6xdx)dx + (5dy)dx + (5dx)dy - (2dy)dy \\ &= 6xdx^2 + 10dxdy - 2dy^2。 \end{aligned}$$

## 二阶条件

运用 $d^2z$ 的概念，我们可以将 $z=f(x,y)$ 极大值的二阶充分条件表述如下：

对于不同时为零的任意 $dx$ 和 $dy$ 值，只要 $d^2z<0$，则存在极大值。　(11.7)

(11.7)的合理性非常类似于单变量情况的 $d^2z$ 条件，它可以通过某曲面的鸟瞰图图 11.4 来解释。令曲面上的点 A〔位于定义域中点 $(x_0,y_0)$ 的正上方〕满足一阶条件(11.3)。则点 $A$ 是极大值的一个有希望的候选点。它是否是极值点实际上还取决于 $A$ 点邻域中曲面的形状。如果沿曲面任意方向的移离 $A$ 点的无穷小运动(参见图 11.4 中的箭头)都会导致 $z$ 的递减，即若对不同时为零的任意 $dx$ 和 $dy$ 值，$dz<0$，则 $A$ 为圆丘之顶。但给定在 $A$ 点 $dz=0$，在 $A$ 邻域中其它点条件 $dz<0$ 意味着规定 $dz$ 是递减的，即对不同时为零的任意 $dx$ 和 $dy$ 值，$d^2z\equiv \mathrm{d}(dz)<0$。因而(11.7)构成了将 $z$ 的稳定值确定为极大值的充分条件。类似的推理还可以证明，将稳定值确定为 $z=f(x,y)$ 的极小值的二阶充分条件为：

对于不同时为零的任意 $dx$ 和 $dy$ 值，

$$d^2z<0。\tag{11.8}$$

316　(11.7)和(11.8)仅是充分条件，但不是必要条件的原因在于，在极大值或极小值，$d^2z$ 也有可能取零值。因此，二阶必要条件必须用弱不等式表述如下：

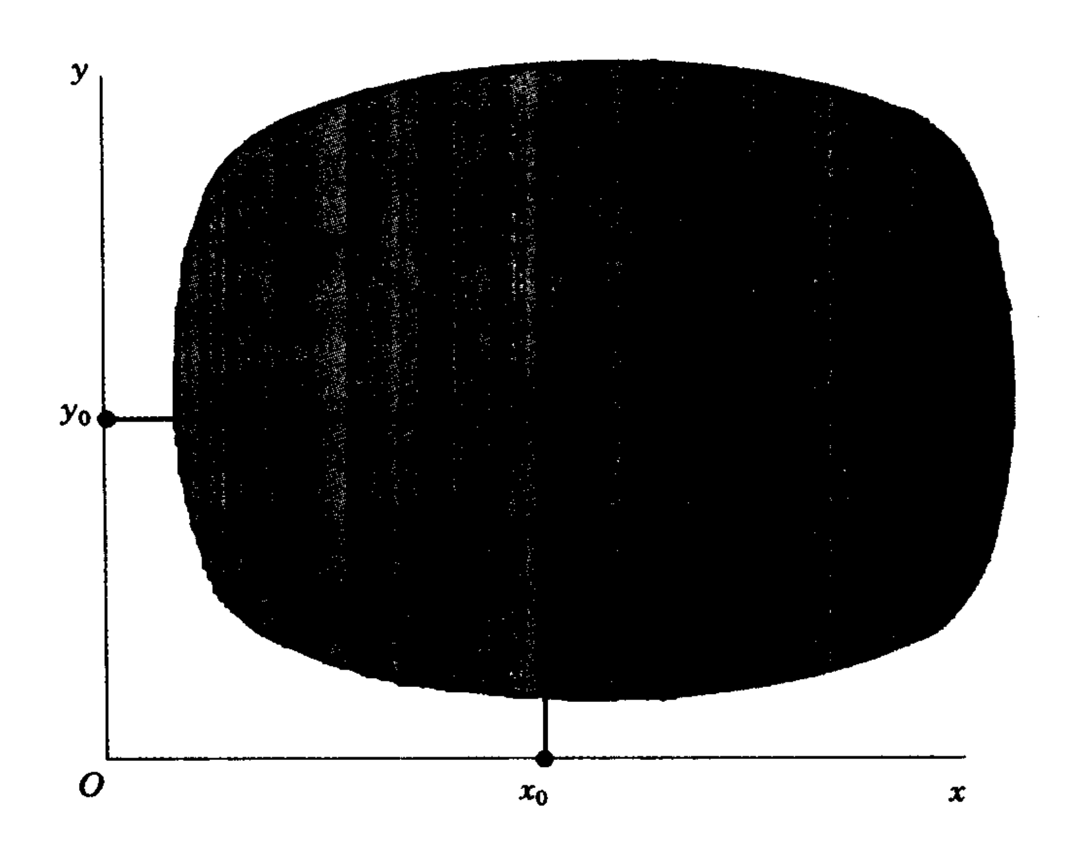



图 11.4

$$
\left.\begin{array}{l} z\text{ 的极大值}: d^2 z \leqslant 0 \\ z\text{ 的极小值}: d^2 z \geqslant 0 \end{array}\right\} \text{对于不同时为零的任意 } dx \text{ 和 } dy \text{ 值。}
$$

但在下面,我们将把更多的注意力放在二阶充分条件上。

为便于运算,二阶微分条件可以转换成等价的二阶导数条件。在两变量情况下,(11.6)表明要对二阶偏导数 $f_{xx}$、$f_{xy}$、$f_{yy}$施加一些限制,实际变换需要了解二次型,我们下一节要对其进行讨论。但我们可以在这里先介绍主要结果:对于不同时为零的任意 $dx$ 和 $dy$ 值,

$$
d^2 z \begin{cases} < 0 & \text{当且仅当} \quad f_{xx} < 0;\ f_{yy} < 0; \quad \text{和} \quad f_{xx} f_{yy} > f_{xy}^2 \\ > 0 & \text{当且仅当} \quad f_{xx} > 0;\ f_{yy} > 0; \quad \text{和} \quad f_{xx} f_{yy} > f_{xy}^2 \end{cases}
$$

注意,$d^2 z$ 的符号不仅取决于 $f_{xx}$ 和 $f_{yy}$ 的符号〔它们的符号与 $A$ 点

(图 11.4)附近由 $T_x$(东西)和 $T_y$(南北)所示的两个基本方向的曲面图形有关],而且与交叉偏导数 $f_{xy}$有关。后一个偏导数所起的作用是保证所研究的曲面能够产生(二维)横截面;这些横截面不仅在两个基本方向(东西和南北),而且在所有其它可能的方向(如东北—西南等)都具有相同类型的图形(山峰或低谷,如本例所示)。

上述结果,同一阶条件(11.5)一起,使我们得到表 11.1。可以理解,表中所有二阶偏导数都在 $f_x = f_y = 0$ 的稳定点计值。还应强调的是,极值的二阶充分条件并**不是必要条件**。特别地,若稳定值具有 $f_{xx}f_{yy} = f_{xy}^2$ 的特征,违背二阶条件,稳定值仍有可能是极值。另一方面,在另一种违背二阶条件的情况下,稳定值具有 $f_{xx}f_{yy} < f_{xy}^2$ 的特征,我们可以确认此点为鞍点,因为在此情况下,$d^2z$ 的符号是不确定的(对于某些 $dx$ 和 $dy$ 值为正,对于另一些则为负。)。

317

**表 11.1　相对极值的条件:$z = f(x, y)$**

| 条件 | 极大值 | 极小值 |
|---|---|---|
| 一阶必要条件 | $f_x = f_y = 0$ | $f_x = f_y = 0$ |
| 二阶充分条件* | $f_{xx}, f_{yy} < 0$ 和 $f_{xx}f_{yy} > f_{xy}^2$ | $f_{xx}, f_{yy} > 0$ 和 $f_{xx}f_{yy} > f_{xy}^2$ |

* 仅在一阶必要条件满足之后方可应用。

**例 4**　求函数

$z = 8x^2 + 2xy - 3x^2 + y^2 + 1$ 的极值。首先,我们求出所有一阶、二阶导数

$$f_x = 24x^2 + 2y - 6x \quad f_y = 2x + 2y$$

$$f_{xx}=48x-6 \quad f_{yy}=2 \quad f_{xy}=2$$

一阶条件要求方程在瞬间满足 $f_x=0, f_y=0$，即

$$24x^2+2y-6x=0$$
$$2y+2x=0$$

第二个方程意味着 $y=-x$，将其代入第一个方程，我们得到 $24x^2-8x=0$，由其可得两个解

$$\bar{x}_1=0 \qquad [\text{意味着} \quad \bar{y}_1=-\bar{x}_1=0]$$

$$\bar{x}_2=\frac{1}{3} \qquad [\text{意味着} \quad \bar{y}_2=-\frac{1}{3}]$$

为应用二阶条件，我们注意到，当

$$\bar{x}_1=\bar{y}_1=0$$

时，$f_{xx}=-6$，而 $f_{yy}=2$，所以 $f_{xx}f_{yy}$ 为负，必然小于平方值 $f_{xy}^2$。它 318 不满足二阶条件。当然，$f_{xx}$ 和 $f_{yy}$ 具有相反符号这一事实表明，所研究的曲面在一个方向向上卷曲，在另一个方向向下卷曲，因而产生了鞍点。

另一个解的情况如何呢？当在 $\bar{x}_2=1/3$ 计值时，我们求得 $f_{xx}=10$，加之 $f_{yy}=f_{xy}=2$，满足了极小值二阶充分条件的三部分内容。因此，通过令给定方程中的 $x=1/3, y=-1/3$，我们可以得到 $z$ 的极小值 $\bar{z}=23/27$。在此例中，仅存在一个相对极值（极小值），它可以用有序三元组来表示

$$(\bar{x}, \bar{y}, \bar{z})=\left(\frac{1}{3}, \frac{-1}{3}, \frac{23}{27}\right)$$

**例 5** 求函数

$$z=x+2ey-e^x-e^{2y}$$

的相对极值。此函数的有关导数为

$$f_x = 1 - e^x \quad f_y = 2e - 2e^{2y}$$

$$f_{xx} = -e^x \quad f_{yy} = -4e^{2y} \quad f_{xy} = 0$$

为满足必要条件，必须有

$$1 - e^x = 0$$

$$2e - 2e^{2y} = 0$$

它只有一个解，即 $\bar{x} = 0, \bar{y} = 1/2$。为确定对应于此解的 $z$ 值是否是极值，我们在 $x = 0, y = 1/2$ 处计算二阶导数的值，求得 $f_{xx} = -1$，$f_{yy} = -4e$，$f_{xy} = 0$。因 $f_{xx}, f_{yy}$ 均为负，且因 $(-1)(-4e) > 0$，所以我们可以得出结论：所研究的 $z$ 值，即

$$\bar{z} = 0 + e - e^0 - e^1 = -1$$

为函数的极大值。这个在给定曲面上的极大值点可以有序三元组 $(\bar{x}, \bar{y}, \bar{z}) = (0, \frac{1}{2}, -1)$ 来表示。

## 练习 11.2

---

运用表 11.1 分别求下列四个函数的极值，并确定其为极大值还是极小值。

1 $z = x^2 + xy + 2y^2 + 3$

2 $z = -x^2 + xy - y^2 + 2x + y$

319 3 $z = ax^2 + by^2 + c$，考察下面三种情况：

(a) $a > 0, b > 0$ (b) $a < 0,\ b < 0$

(c) $a$ 与 $b$ 符号相反

4 $z = e^{2x} - 2x + 2y^2 + 3$

5 考察函数 $z = (x-2)^4 + (y-3)^4$。

(a) 用直观推理确定 $z$ 在 $\bar{x} = 2, \bar{y} = 3$ 处达到极小值，$\bar{z} = 0$。

(*b*)表 11.1 中的一阶必要条件得到满足了吗?

(*c*)表 11.1 中的二阶充分条件得到满足了吗?

(*d*)求 $d^2z$ 的值。它满足极小值的二阶必要条件(11.9)吗?

---

## 11.3 二次型——偏离主题的讨论

在 $d^2z$ 表达式(11.6)最后一行,给出了通称为二次型的一个例子。对于二次型,存在现成的标准判断对于不同时为零的任意 $dx$ 和 $dy$,其符号恒为正、负,还是非正或者非负。因为极值的二阶条件直接依赖于 $d^2z$ 的符号,所以这些判别标准有重要意义。

首先,我们把型定义为各项具有相同次数的多项式。我们前面遇到的多项式都被限定在单一变量的情况:$a_0 + a_1x + a_2x + \cdots + a_nx$。当包含更多变量时,多项式的每一项可能含有一个或几个变量,每个变量都自乘至非负整数幂,如 $3x + 4x^2y^3 - 2yz$ 等。在多项式中的每一项都有相同次数,即每一项的指数和相同的特殊情况下,这种多项式被称作型。例如,$4x - 9y + z$ 是一个三变量的线性型,因为其中的每一项都是一次的。另一方面,$4x^2 - xy + 3y^2$,其中每一项均为二次(整数指数和为 2),所以构成了一个两个变量的二次型。我们也可能遇到三变量的二次型,如 $x^2 + 2xy - yw + 7w^2$,或者 n 个变量的二次型等。

### 二次型的二阶全微分

若我们将(11.6)中的微分 $dx$ 和 $dy$ 视为变量,将偏导数视为系数,即若令

$$\begin{aligned} &u \equiv dx \qquad v \equiv dy \\ &a \equiv f_{xx} \qquad b \equiv f_{yy} \qquad h \equiv f_{xy}[=f_{yx}] \end{aligned} \tag{11.10}$$

则二阶全导数

$$q = au^2 + 2huv + bv^2 \tag{11.6'}$$

注意,在此二次型中,$dx \equiv u$, $dy \equiv v$ 充当变量的角色,而二次偏 320 导数则被视为常数,这与我们微分 $dz$ 以得到 $d^2z$ 时的情形恰恰相反。角色颠倒的原因在于我们现在研究的问题的性质发生了变化。极值的二阶充分条件规定,无论 $dx$ 和 $dy$ 可能取何值,只要它们不同时为零,对于极小值,$d^2z$ 必然为正;对于极大值,$d^2z$ 必定为负。因此很明显,在现在的内容中,$dx$ 和 $dy$ 必须被视为**变量**。而二阶偏导数,在我们考察可能为极值的点上取具体值,所以可以视为**常数**。

这样,主要问题变成:当允许(11.6′)中的 $u$ 和 $v$ 可以取任意值时,为了得到确定符号的 $q$,应对 $a$、$b$ 和 $h$ 施加何种限制?

## 正定与负定

就一个术语而言,我们可将二次型 $q$ 表述成:

$$\text{若 } q \text{ 恒为} \begin{cases} \text{正} & (>0) \\ \text{非负} & (\geqslant 0) \\ \text{非正} & (\leqslant 0) \\ \text{负} & (<0) \end{cases}, \text{则 } q \text{ 为} \begin{cases} \text{正定} \\ \text{半正定} \\ \text{半负定} \\ \text{负定} \end{cases}$$

不管二次型中的变量取何值,只要不全为零。另一方面,当变量取不同值时,$q$ 的符号发生了变化,则称 $q$ 是**不定的**。很明显,$q = d^2z$ 为正定或负定的情况是分别与极小值或极大值的二阶充分条件联系在一起的。而**半定**的情况则是与二阶**必要**条件相联系的。

当 $q=d^2z$ 为不定时,则出现了鞍点的征兆。

## 二次型有定符号的行列式检验

广泛使用的 $q$ 的有定符号检验,需要考察某行列式的符号。这种方法恰好可以更容易地应用于正定和负定(相对于半定而言)。即它可以更方便地应用于二阶充分条件(与必要条件相比较)。所以,我们这里仅限于讨论充分条件。[①]

对于两变量的情况,$q$ 的确定符号的行列式条件是相对容易推导的。首先我们看到,(11.6′)中的第一项、第三项与变量 $u$ 和 $v$ 的值无关,因这些变量是以平方形式出现的。因此,通过限制 $a$ 321 和 $b$ 的符号,容易单独确定这些项正定和负定的条件。麻烦的是中间那项。但若我们能将多项式变换成变量 $u$ 和 $v$ 仅在平方项中出现的表达式,那么 $q$ 的符号的确定就容易了。

完成这一任务的方法是配方。通过在(11.6′)中加上并减去 $h^2v^2/a$,我们可将二次型改写成如下形式:

$$
\begin{aligned}
q &= au^2 + 2huv + \frac{h^2}{a}v^2 + bv^2 - \frac{h^2}{a}v^2 \\
&= a\left(u^2 + \frac{2h}{a}uv + \frac{h^2}{a^2}v^2\right) + \left(b - \frac{h^2}{a}\right)v^2 \\
&= a\left(u + \frac{h}{a}v\right)^2 + \frac{ab - h^2}{a}(v^2)
\end{aligned}
$$

注意,变量 $u$ 和 $v$ 仅在平方项中出现,我们完全可以根据系数 $a$、$b$ 和 $h$ 的如下值确定 $q$ 的符号:

---

① 关于二阶必要条件行列式检验的讨论,请参阅高山昭《数理经济学》,德瑞登出版公司,1974 年版,pp. 118—120。

$$\text{当且仅当}\begin{cases} a>0 \\ a<0 \end{cases} \text{且 } ab-h^2>0, \qquad q\text{ 为}\begin{cases}\text{正定}\\ \text{负定}\end{cases}$$

还要注意(1)在两种情况下,$ab-h^2$ 均应为正;(2)作为 $ab-h^2$ 为正的前提条件,积 $ab$ 必须为正,因为它必须大于平方项 $h^2$。因此,上述条件必然意味着 $a$ 与 $b$ 必须取相同的代数符号。

刚才推导出来的条件可用行列式简洁表示。首先,我们注意到,(11.6′)中的二次型可重排成如下对称的方形:

$$\begin{aligned} q = a(u^2) &+ h(uv) \\ + h(vu) &+ b(v^2) \end{aligned}$$

将平方项置于对角位置,将 $2huv$ 拆分成两个相等的部分,也分别置于对角位置。现在系数构成一个对称矩阵:$a$ 和 $b$ 在主对角线上,$h$ 在非对角线上。由此看来,二次型可以视为由下列矩阵乘法得到的 $1\times1$ 矩阵(标量):

$$q=\begin{bmatrix} u & v \end{bmatrix}\begin{bmatrix} a & h \\ h & b \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix}$$

$2\times2$ 系数矩阵的行列式 $\begin{vmatrix} a & h \\ h & b \end{vmatrix}$,可以称作二次型 $q$ 的判别式,我们以 $|D|$ 来表示。它为(11.11)中的判别准则用行列式表示提322供了线索,因为此准则也可以表示成:

$$\text{当且仅当}\begin{Bmatrix} |a|>0 \\ |a|<0 \end{Bmatrix} \text{且} \begin{vmatrix} a & h \\ h & b \end{vmatrix}>0, \quad q\text{ 为}\begin{cases}\text{正定}\\ \text{负定}\end{cases} \qquad (11.11')$$

行列式 $|a|=a$ 是 $|D|$ 的一个子行列式,它由主对角线的第一个元素组成,称作 $|D|$ 的第一个主子式。行列式 $\begin{vmatrix} a & h \\ h & b \end{vmatrix}$ 也可以视为 $|D|$ 的子行列式,因为它包含了主对角线的第一个和第二个元素。我们可称其为 $|D|$ 的第二个主子式。在本例中,仅有两个

主子式，其符号将用于确定 $q$ 是正定还是负定。

通过(11.10)，当(11.11′)各项以二阶全微分 $d^2z$ 的各项表示时，我们有：

$$\text{当且仅当}\begin{Bmatrix} f_{xx}>0 \\ f_{xx}<0 \end{Bmatrix} \quad \text{和} \quad \begin{vmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{vmatrix} = f_{xx}f_{yy} - f_{xy}^2 > 0$$

$$d^2z \text{ 为}\begin{cases} \text{正定} \\ \text{负定} \end{cases}$$

回想一下，上式最后一个不等式意味着 $f_{xx}$ 和 $f_{yy}$ 应取**相同**的符号，这实际上是在表 11.1 中所列的二阶充分条件。

一般而言，二次型

$$q = au^2 + 2huv + 6v^2$$

的行列式为对称行列式 $\begin{vmatrix} a & h \\ h & b \end{vmatrix}$。在具体二次型

$$d^2z = f_{xx}dx^2 + 2f_{xy}dxdy + f_{yy}dy^2$$

中，其判别式是以二阶偏导数为元素的行列式。此行列式被称为**海塞行列式**(或简称**海塞**)。在两变量情况下，海塞行列式为

$$|H| = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix}$$

根据扬氏定理($f_{xy} = f_{yx}$)，此行列式是对称的——作为判别式它也应当如此。读者应将海塞行列式与 7.6 节讨论的雅可比行列式区别开来。

**例 1** 判断 $q = 5u^2 + 3uv + 2v^2$ 为正定还是负定。$q$ 的判别式为 $\begin{vmatrix} 5 & 1.5 \\ 1.5 & 2 \end{vmatrix}$，其主子式为

$$5 > 0 \quad \text{和} \quad \begin{vmatrix} 5 & 1.5 \\ 1.5 & 2 \end{vmatrix} = 7.75 > 0$$

因此 $q$ 为正定的二次型。

323　**例2**　在函数 $z=f(x,y)$ 的某点，已知 $f_{xx}=-2, f_{xy}=1, f_{yy}=-1$，无论 $dx$ 与 $dy$ 取何值，$d^2z$ 在该点有确定符号吗？在本例中，二次型 $d^2z$ 的判别式 $\begin{vmatrix} -2 & 1 \\ 1 & -1 \end{vmatrix}$，具有主子式

$$-2<0 \quad 和 \quad \begin{vmatrix} -2 & 1 \\ 1 & -1 \end{vmatrix}=1>0$$

因此，$d^2z$ 是负定的。

## 三变量二次型

关于三变量二次型，也可得到类似条件吗？

具有三个变量 $u_1, u_2, u_3$ 的二次型一般可表示成

$$\begin{aligned} q(u_1,u_2,u_3) &= d_{11}(u_1^2)+d_{12}(u_1u_2)+d_{13}(u_1u_3) \\ &\quad + d_{21}(u_2u_1)+d_{22}(u_2^2)+d_{23}(u_2u_3) \\ &\quad + d_{31}(u_3u_1)+d_{32}(u_3u_2)+d_{33}(u_3^2) \\ &= \sum_{i=1}^{3}\sum_{j=1}^{3} d_{ij}u_iu_j \end{aligned} \tag{11.12}$$

其中双 $\sum$（双重求和）符号表示指标 $i$ 和 $j$ 均可取值1、2、3。因此双重求和表达式等价于上面所表示的 $3\times 3$ 阵列。顺便提一句，尽管我们已将每对系数写成 $(d_{12}, d_{21})$ 或 $(d_{23}, d_{32})$ 这种形式，似乎每对中的两个数是不同的，但二次型的方形阵列总是可以看作对称的。因为如果二次型包含的变量 $u_1$ 和 $u_2$ 的项，比如说恰好是 $12u_1u_2$，我们总可以令 $d_{12}=d_{21}=6$，使得 $d_{12}u_1u_2=d_{21}u_2u_1$。应用类似的办法还可以使其它非对角元素对称。

实际上，三变量二次型还可以表示成三个矩阵之积：

$$q(u_1,u_2,u_3) = [u_1\ u_2\ u_3]\begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \equiv u'Du \tag{11.12'}$$

同在两变量情况一样，第一个矩阵(行向量)与第三个矩阵(列向量)仅列出了变量，中间矩阵($D$)是源于二次型(11.12)的方阵阵列的对称系数矩阵。但这里，由判别式共可形成三个主子式，即 324

$$|D_1| \equiv d_{11} \qquad |D_2| \equiv \begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix}$$

$$|D_3| \equiv \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix}$$

其中$|D_i|$表示判别式$|D|$的第 $i$ 个主子式。[①] 可以证明，对这些主子式的符号加以限制，也可用来表示正定和负定的条件。

运用刚熟悉的配方法，(11.12)中的二次型可以变换成三个变量仅以平方形式出现的表达式。具体地，回想 $d_{12}=d_{21}$ 等，我们有

$$q = d_{11}\left(u_1+\frac{d_{12}}{d_{11}}u_2+\frac{d_{13}}{d_{11}}u_3\right)^2 + \frac{d_{11}d_{22}-d_{12}^2}{d_{11}}\left(u_2+\frac{d_{11}d_{23}-d_{12}d_{13}}{d_{11}d_{22}-d_{12}^2}u_3\right)^2$$

---

① 迄今为止，我们一直把第 $i$ 个主子式$|D_i|$看成是由包含$|D|$的前 $i$ 个主对角元素所形成的子行列式。但因子式的概念就是从原行列式删去某些行和列而得到的子行列式，所以也可把第 $i$ 个主子式看作删去最后$(n-i)$行和列后所得到的子行列式。

$$+\frac{d_{11}d_{22}d_{33}-d_{11}d_{23}^2-d_{22}d_{13}^2-d_{33}d_{12}^2+2d_{12}d_{13}d_{23}}{d_{11}d_{22}-d_{12}^2}(u_3)^2$$

对于任意不同时为零的 $u_1$、$u_2$、$u_3$，当且仅当三个平方表达式的系数均为正(负)时，平方和为正(负)。但三个系数按给定顺序，可以按如下三个主子式表示：

$$|D_1| \quad \frac{|D_2|}{|D_1|} \quad \frac{|D_3|}{|D_2|}$$

因此，**正定**的充要条件包括三个方面

$$|D_1|>0$$

$$|D_2|>0 \qquad [\text{已知} \quad |D_1|>0]$$

$$|D_3|>0 \qquad [\text{已知} \quad |D_2|>0]$$

换句话说，即三个主子式必须都为正。而**负定**的充要条件则为

$$|D_1|<0$$

$$|D_2|>0 \qquad [\text{已知} \quad |D_1|<0]$$

$$|D_3|<0 \qquad [\text{已知} \quad |D_2|>0]$$

即三个主子式必须按规定的方式交替改变符号。

325 **例 3** 确定

$$q=u_1^2+6u_2^2+3u_3^2-2u_1u_2-4u_2u_3$$

是正定还是负定，$q$ 的判别式是：

$$\begin{vmatrix} 1 & -1 & 0 \\ -1 & 6 & -2 \\ 0 & -2 & 3 \end{vmatrix}$$

其主子式如下：

$$1>0 \quad \begin{vmatrix} 1 & -1 \\ -1 & 6 \end{vmatrix}=5>0 \quad \text{和} \quad \begin{vmatrix} 1 & -1 & 0 \\ -1 & 6 & -2 \\ 0 & -2 & 3 \end{vmatrix}=11>0$$

因此,二次型为正定的。

**例4** 确定

$$q = 2u^2 + 3v^2 - w^2 + 6uv - 8uw - 2vw$$

是正定还是负定。判别式可以写成

$$\begin{vmatrix} 2 & 3 & -4 \\ 3 & 3 & -1 \\ -4 & -1 & -1 \end{vmatrix},$$

我们求得其第一个主子式为 $2>0$,但第二个主子式 $\begin{vmatrix} 2 & 3 \\ 3 & 3 \end{vmatrix} = -3 < 0$。这既违背了正定的条件,也违背了负定的条件,所以此二次型既非正定,也非负定。

## *n* -变量二次型

把上述结果推广到 $n$ 变量情况,我们可以这样表述(不作证明):对于二次型

$$q(u_1, u_2, \cdots, u_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} u_i u_j \qquad [\text{其中 } d_{ij} = d_{ji}]$$

$$= \underset{(1\times n)}{u'} \quad \underset{(n\times n)}{D} \quad \underset{(n\times 1)}{u} \qquad [\text{参见}(11.12')]$$

正定的充要条件为 $|D|$ 的主子式,即

$$|D_1| \equiv d_{11} \qquad |D_2| \equiv \begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix} \qquad \cdots$$

$$|D_n| \equiv \begin{vmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{vmatrix}$$

均为正。相应地，负定的充要条件是：主子式按如下方式交替改变符号：

$$|D_1|<0 \qquad |D_2|>0 \qquad |D_3|<0\ (\text{等})$$

使得所有奇数主子式为负，所有偶数主子式为正。第 $n$ 个主子式 326 $|D_n|=|D|$，若 $n$ 为偶数，则为正；若 $n$ 为奇数，则为负。此结论可通过不等式 $(-1)^n|D_n|>0$ 简洁地表示。

## 二次型有定符号的特征根检验

对二次形 $u'Du$ 的有定符号，除了上面介绍的行列式检验外，还有另一种利用所谓矩阵 $D$ 的"特征根"概念来进行检验的方法。特征根这个概念出现在具有如下性质的问题中：给定一个 $n\times n$ 矩阵 $D$，我们能否求得标量 $r$ 和 $n\times 1$ 向量 $x\neq 0$，使得矩阵方程

$$\underset{(n\times n)}{D}\ \underset{(n\times 1)}{x}=r\ \underset{(n\times 1)}{x} \tag{11.13}$$

成立？若能，则标量 $r$ 称作矩阵 $D$ 的**特征根**，向量 $x$ 称作该矩阵的**特征向量**。[①]

矩阵方程 $Dx=rx$ 可重写成 $Dx-rIx=0$ 或者

$$(D-rI)x=0 \qquad \text{其中 0 是 } n\times 1 \tag{11.13'}$$

当然，此式表示 $n$ 个齐次线性方程的方程组。因为我们要求出 $x$ 的非平凡解，所以要求系数矩阵 $(D-rI)$（称作矩阵 $D$ 的**特征矩阵**）是奇异矩阵。换言之，其行列式必为零

① 特征根还可称为**本征根**，或本征值，特征向量也可称作本征向量。

$$
|D - rI| = \begin{vmatrix} d_{11} - r & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} - r & \cdots & d_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} - r \end{vmatrix} = 0 \tag{11.14}
$$

方程(11.14)被称作矩阵 $D$ 的**特征方程**。因为行列式 $|D - rI|$ 的拉普拉斯展开会得到变量 $r$ 的 $n$ 次多项式，所以(11.14)实际上是一个 $n$ 次多项式方程。于是它共有 $n$ 个根$(r_1, \cdots, r_n)$，每一个根都是特征根。若 $D$ 是对称的，像二次型中的情形一样，则其特征根将总是实数，但他们可能为正，也可能为负，或者为零。

由于这些 $r$ 值均使行列式 $|D - rI|$ 等于零，因此，将上述任何一个 $r$ 值，比如 $r_i$ 代入方程组(11.13′)，都将产生相应的向量 $x|_{r=r_i}$。更确切地说，作为齐次方程组，对应于特征根 $r_i$，会产生无数个向量。但我们可用**正规化**办法(后面解释)，从无数个向量中选择特定的向量作为对应于 $r_i$ 的特征向量；此向量可以 $v_i$ 表示。因有 $n$ 个特征根，所以对应的特征向量应共同 $n$ 个。 327

**例 5** 求矩阵

$$
\begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}
$$

的特征向量和特征根。将已知矩阵 $D$ 代入(11.14)，得到方程

$$
\begin{vmatrix} 2 - r & 2 \\ 2 & -1 - r \end{vmatrix} = r^2 - r - 6 = 0
$$

其根为 $r_1 = 3, r_2 = -2$。利用第一个根，矩阵方程(11.13′)取如下形式：

$$\begin{bmatrix} 2-3 & 2 \\ 2 & -1-3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

正如我们根据(11.14)所预料的那样，系数矩阵的两行线性相关，方程有无穷多个解。它可以方程 $x_1 = 2x_2$ 来表示，为得出唯一解，我们施加限制 $x_1{}^2 + x_2{}^2 = 1$ 以使解正规化。[①] 则因

$$x_1^2 + x_2^2 = (2x_2)^2 + x_2^2 = 5x_2^2 = 1$$

取正的平方根，我们有 $x_2 = 1/\sqrt{5}$，和 $x_1 = 2x_2 = 2/\sqrt{5}$。因此第一个特征向量为

$$v_1 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$$

类似地，在(11.13′)中运用第二个特征根，得到方程

$$\begin{bmatrix} 2-(-2) & 2 \\ 2 & -1-(-2) \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

方程有解 $x_1 = -\frac{1}{2}x_2$。通过正规化，我们可得到

$$x_1^2 + x_2^2 = (-\frac{1}{2}x_2)^2 + x_2^2 = \frac{5}{4}x_2^2 = 1$$

它有解 $x_2 = 2/\sqrt{5}$ 和 $x_1 = -1/\sqrt{5}$。所以第二个特征向量为：

$$v_2 = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

以此方式得到的特征向量集有两个重要性质：首先，标量积 $v_i{}'v$ 328 $(i = 1, 2, \cdots, n)$ 必定等于 1，因为

① 更一般地，对于 $n$ – 变量的情况，我们要求 $\sum_{i=1}^{n} x_i{}^2 = 1$。

$$v_i'v_i = [x_1\ x_2\ \cdots\ x_n]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} x_i^2 = 1 \quad [\text{通过正规化}]$$

第二,标量积 $v_i'v_j$(其中 $i \neq j$)总为零。[1] 因此,我们可以写成

$$v_i'v_i = 1 \qquad \text{和} \qquad v_i'v_j = 0 \qquad (i \neq j) \qquad (11.15)$$

后面会看到,这个性质很有用处。就术语而言,当两个向量的标量积为零时,可称它们互为正交(垂直)。[2] 因此,矩阵 $D$ 的每对特征向量必然为正交的。另一个性质 $v_i'v_i = 1$ 表示正规化。这两个性质合起来解释了特征向量($v_1, \cdots, v_n$)被称作一组标准正交向量的原因。读者可以试证例 5 中的两个特征向量为标准正交向量。

现在,我们开始解释矩阵 $D$ 的特征根和特征向量如何用于确

---

① 为证明这一点,由(11.13),我们可以写出 $Dv_j = r_jv_j$ 和 $Dv_i = r_iv_i$。以适当的行向量左乘上述方程两边,我们有

$$v_i'Dv_j = v_i'r_jv_j = r_jv_i'v_j \qquad [r_j\ \text{是标量}]$$

$$v_j'Dv_i = v_j'r_iv_i = r_iv_j'v_i = r_iv_i'v_j \qquad [v_i'v_j = v_j'v_i]$$

因 $v_i'Dv_j$ 和 $v_j'Dv_i$ 均为 $1\times 1$ 矩阵,且二者可互换(回忆一下,因 $D$ 是对称的,所以 $D' = D$),所以,它们必然表示相同的标量。由此可知,上面两个方程最右边的表达式相等。因此,将其相减,有

$$(r_j - r_i)v_i'v_j = 0$$

现在,若 $r_j \neq r_i$(不同的根),为使方程成立,$v_i'v_j$ 必须为零,这便证明了我们的结论。而且,若 $r_j = r_i$(重根),则总可以找到两个线性无关的正规化的向量,满足 $v_i'v_j = 0$。因此,我们可以作如下表述:一般地,当 $i \neq j$ 时,$v_i'v_j = 0$。

② 作为这一问题的简单描述,考虑二维空间中的两个单位向量 $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 和 $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$。这两个向量分别位于两个轴上,因此是垂直的。同时,我们确实求得 $e_1'e_2 = e_2'e_1 = 0$。

定二次型$u'Du$的有定符号。这种思想的实质依然是把$u'Du$(它不仅包含平方项$u_1^2, \cdots, u_n^2$,而且包含$u_1u_2, u_2u_3$这样的交叉积)转换成仅含平方项的形式,因此,方法也与前面推导行列式检验时所用的配方法类似。但在这里变换还有新的特点:每个平方项均有一个特征根系数,所以$n$个根的符号足以确定二次型的有定符号。

329 实现上述目的的变换如下:令特征向量$v_1, \cdots, v_n$构成特征矩阵$T$的列:

$$\underset{(n\times n)}{T} = [v_1 \quad v_2 \quad \cdots \quad v_n]$$

则应用变换$\underset{(n\times 1)}{u} = \underset{(n\times n)}{T}\underset{(n\times 1)}{y}$ 于二次型$u'Du$:

$$\begin{aligned} u'Du &= (Ty)'D(Ty) = y'T'DTy \quad [\text{由}(4.11)] \\ &= y'Ry \quad \text{其中 } R \equiv T'DT \end{aligned}$$

因此,变量$u_i$的初始二次型现在变换成为另一个变量$y_i$的二次型。因变量$u_i$和$y_i$取相同的值域,因此变换并不影响二次型的有定符号。这样,我们便可以考察二次型$y'Ry$的符号了。我们对后一个二次型感兴趣的是,矩阵$R$是一个对角阵,矩阵$D$的根$r_1$,…,$r_n$位于对角线上其它各处均为零,从而我们事实上有

$$u'Du = y'Ry = [y_1\ y_2 \cdots y_n]\begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & r_n \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$= r_1y_1^2 + r_2y_2^2 + \cdots + r_ny_n^2 \tag{11.16}$$

它是一个仅含平方项的表达式。因此,变换$R \equiv T'DT$给我们提

供了一种将对称矩阵$D$对角化成特定对角阵$R$的方法。

**例6** 验证在例5中给出的矩阵$\begin{bmatrix}2 & 2\\2 & -1\end{bmatrix}$可以对角化为矩阵$\begin{bmatrix}r_1 & 0\\0 & r_2\end{bmatrix}=\begin{bmatrix}3 & 0\\0 & -2\end{bmatrix}$。根据例5中求出的特征向量,变换矩阵$T$应为

$$T=[v_1 \quad v_2]=\begin{bmatrix}2/\sqrt{5} & -1/\sqrt{5}\\1/\sqrt{5} & 2/\sqrt{5}\end{bmatrix}$$

因此,我们可以写成

$$R\equiv T'DT=\begin{bmatrix}\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}}\\-\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}}\end{bmatrix}\begin{bmatrix}2 & 2\\2 & -1\end{bmatrix}\begin{bmatrix}\frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}}\\\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}}\end{bmatrix}=\begin{bmatrix}3 & 0\\0 & -2\end{bmatrix}$$

它完全验证了对角化过程。

为证明(11.16)中的对角化结果,我们部分写出矩阵$R$如下: 330

$$R\equiv T'DT=\begin{bmatrix}v'_1\\v'_2\\\vdots\\v'_n\end{bmatrix}D[v_1\ v_2\cdots\ v_n]$$

很容易验证,$D[v_1\ v_2\cdots\ v_n]$可重写成为$[Dv_1\ Dv_2\cdots\ Dv_n]$。此外,由(11.13),我们还可以进一步将其改写成$[r_1v_1\ r_2v_2\cdots\ r_nv_n]$。因此,我们看到

$$R=\begin{bmatrix}v'_1\\v'_2\\\vdots\\v'_n\end{bmatrix}[r_1v_1 \quad r_2v_2 \quad \cdots \quad r_nv_n]$$

$$
= \begin{bmatrix} r_1 v_1' v_1 & r_2 v_1' v_2 & \cdots & r_n v_1' v_n \\ r_1 v_2' v_1 & r_2 v_2' v_2 & \cdots & r_n v_2' v_n \\ \cdots & \cdots & \cdots & \cdots \\ r_1 v_n' v_1 & r_2 v_n' v_2 & \cdots & r_n v_n' v_n \end{bmatrix}
$$

$$
= \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & r_n \end{bmatrix} \qquad [\text{由}(11.15)]
$$

这恰好是我们想证明的。

根据(11.16)中的结果,我们可以正式地将二次型有定符号的特征根检验表述如下:

*a* 当且仅当 $D$ 的每个特征根为正(负),则 $q = u'Du$ 为正(负)定。

*b* 当且仅当 $D$ 的所有特征根为非负(非正),且至少有一个特征根为零,则 $q = u'Du$ 为半正(负)定。

*c* 当且仅当 $D$ 的某些特征根为正,而另一些为负时,$q = u'Du$ 是不定的。

注意,在应用这一检验时,我们所需要的就是特征根;除非我们希望求出变换矩阵 $T$,否则,我们不需要特征向量。还要注意,不同于前面介绍的行列式检验,这种检验方法使我们能够同时检验二阶必要条件(上面的 $b$ 部分)和二阶充分条件($a$ 部分)。但这种检验方法也具有一个缺点。当矩阵 $D$ 阶数高时,很难解多项式方程(11.14)以得到这种检验所需要的特征根。在这种情况下,行列式检验可能是一种更好的方法。

## 练习 11.3



1 直接运用矩阵乘法，将下面的矩阵积表示为二次型：

(a) $[u\ v]\begin{bmatrix}4 & 2\\ 2 & 3\end{bmatrix}\begin{bmatrix}u\\ v\end{bmatrix}$　　(c) $[x\ y]\begin{bmatrix}5 & 2\\ 4 & 0\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}$

(b) $[u\ v]\begin{bmatrix}-2 & 3\\ 1 & -4\end{bmatrix}\begin{bmatrix}u\\ v\end{bmatrix}$　　(d) $[dx\ dy]\begin{bmatrix}f_{xx} & f_{xy}\\ f_{yx} & f_{yy}\end{bmatrix}\begin{bmatrix}dx\\ dy\end{bmatrix}$

2 在上题的 $b$ 和 $c$ 中，系数矩阵对主对角线不是对称的。验证通过取非对角元素的平均值，因而将其分别转换成 $\begin{bmatrix}-2 & 2\\ 2 & -4\end{bmatrix}$ 和 $\begin{bmatrix}5 & 3\\ 3 & 0\end{bmatrix}$，我们得到的二次型同前面是一样的。

3 根据其系数矩阵(对称形式)，运用行列式检验确定练习 11.3－1 的 $a$、$b$、$c$ 中的二次型是正定还是负的。

4 将下面的每个二次型表示成包含对称系数矩阵的矩阵积：

(a) $q = 3u^2 - 4uv + 7v^2$

(b) $q = u^2 + 7uv + 3v^2$

(c) $q = 8uv - u^2 - 31v^2$

(d) $q = 6xy - 5y^2 - 2x^2$

(e) $q = 3u_1^2 - 2u_1u_2 + 4u_1u_3 + 5u_2^2 + 4u_3^2 - 2u_2u_3$

(f) $q = -u^2 + 4uv - 6uw - 4v^2 - 7w^2$

5 由从上题对称系数矩阵得到的判别式，运用行列式检验，判定哪个二次型为正定，哪个二次型为负定。

6 求下列每个矩阵的特征根

(a) $D = \begin{bmatrix}4 & 2\\ 2 & 3\end{bmatrix}$　(b) $E = \begin{bmatrix}-2 & 2\\ 2 & -4\end{bmatrix}$　(c) $F = \begin{bmatrix}5 & 3\\ 3 & 0\end{bmatrix}$

关于二次型 $u'Du$、$u'Eu$、$u'Fu$ 的符号，能得出什么结论？(对照练习 11.3－3，检验你的结论。)

7 求矩阵 $\begin{bmatrix}4 & 2\\ 2 & 1\end{bmatrix}$ 的特征向量。

8 给定二次型 $u'Du$，其中 $D$ 为 $2\times2$ 矩阵，矩阵 $D$ 的特征方程可以写成

$$\begin{vmatrix} d_{11}-r & d_{12} \\ d_{21} & d_{22}-r \end{vmatrix} = 0 \qquad (d_{12}=d_{21})$$

展开此行列式,用平方公式表示此方程的根,并推断下列结论:

(*a*) 在 $r_1$ 和 $r_2$ 中不存在虚数(包含 $\sqrt{-1}$ 的数)。

(*b*) 要有重根,矩阵 $D$ 必须具有这种形式:$\begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}$。

(*c*) 要得到半正定或半负定,二次型的判别式必为零,即 $|D|=0$。

---

332

## 11.4 具有多于两个变量的目标函数

当目标函数中出现 $n>2$ 个选择变量时,尽管我们仍可以谈及 $(n+1)$ 维空间中的超平面,但不可能绘出函数的图形。在这种不能图形化的超平面中,仍可能存在 $(n+1)$ 维的类似的峰顶和谷底。我们如何识别它们呢?

### 极值的一阶条件

我们具体考察一个具有三个选择变量的函数:

$$z=f(x_1,\ x_2,\ x_3)$$

其一阶偏导数 $f_1, f_2, f_3$ 和二阶导数 $f_{ij}(\equiv\partial^2 z/\partial x_i\partial x_j)$, $i=1,2,3$。利用杨氏定理,我们有 $f_{ij}=f_{ji}$。

我们前面的讨论表明,要得到 $z$ 的极大值或极小值,必须使对于不同时为零的任意 $dx_1, dx_2, dx_3$ 值,$dz=0$。因为 $dz$ 的值现在为

$$dz=f_1dx_1+f_2dx_2+f_3dx_3 \tag{11.17}$$

且因 $dx_1$、$dx_2$、$dx_3$ 是不同时为零的任意(无穷小)变化的自变量,

所以确保 $dz=0$ 的唯一方式是 $f_1=f_2=f_3=0$。因此,同两变量情形一样,极值的必要条件仍是所有一阶偏导数均为零。[①]

## 二阶条件

一阶条件的满足给 $z$ 的某些值打上了目标函数稳定值的标记。若在 $z$ 的一个稳定值,我们发现 $d^2z$ 为正定,便可确定此 $z$ 值为极小值。类似地,$d^2z$ 为负定则是此稳定值为极大值的充分条件。这引出了一个问题:在目标函数中存在三个变量时如何表示 $d^2z$ 呢?如何确定它是正定还是负定呢?

对(11.17)的 $dz$ 求微分,可以得到 $d^2z$ 的表达式。在微分 333 过程中,如在(11.6)中一样,我们应将导数 $f_i$ 作为变量,将微分 $dx_i$ 作为常量对待。因此,我们有

$$
\begin{aligned}
d^2z = d(dz) &= \frac{\partial(dz)}{\partial x_1}dx_1 + \frac{\partial(dz)}{\partial x_2}dx_2 + \frac{\partial(dz)}{\partial x_3}dx_3 \\
&= \frac{\partial}{\partial x_1}(f_1dx_1 + f_2dx_2 + f_3dx_3)dx_1 \\
&\quad + \frac{\partial}{\partial x_2}(f_1dx_1 + f_2dx_2 + f_3dx_3)dx_2 \\
&\quad + \frac{\partial}{\partial x_3}(f_1dx_1 + f_2dx_2 + f_3dx_3)dx_3
\end{aligned}
$$

---

① 作为一个特例,注意若我们讨论的恰好是一个由方程 $F(z, x_1, x_2, x_3)=0$ 隐性地定义的方程 $z=f(x_1, x_2, x_3)$。其中

$$f_i \equiv \frac{\partial z}{\partial x_i} = \frac{-\partial F/\partial x_i}{\partial F/\partial z} \quad (i=1,2,3)$$

则一阶条件 $f_1=f_2=f_3=0$ 等价于条件

$$\frac{\partial F}{\partial x_1} = \frac{\partial F}{\partial x_2} = \frac{\partial F}{\partial x_3} = 0$$

因分母 $\partial F/\partial z \neq 0$ 的值没有关系。

$$
\begin{aligned}
&= f_{11}dx_1^2 + f_{12}dx_1dx_2 + f_{13}dx_1dx_3 \\
&\quad + f_{21}dx_2dx_1 + f_{22}dx_2^2 + f_{23}dx_2dx_3 \\
&\quad + f_{31}dx_3dx_1 + f_{32}dx_3dx_2 \\
&\quad + f_{33}dx_3^2
\end{aligned}
\tag{11.18}
$$

这是一个类似于(11.12)的二次型。因而,我们前面掌握的正定和负定的判别标准可直接在这里应用。

在确定 $d^2z$ 是正定或负定时,我们还要像在(11.6′)中一样,把 $dx_i$ 视为可取任意值但不同时为零的变量,而把导数 $f_{ij}$ 看作施加某些限制的系数。(11.18)中的系数产生对称的海塞行列式

$$
|H| = \begin{vmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{vmatrix}
$$

其逐次主子式可以表示为

$$
|H_1| = f_{11} \qquad |H_2| = \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} \qquad |H_3| = |H|
$$

因此,根据正定和负定的行列式判断标准,我们可将 $z$ 的极值的二阶充分条件表述如下:

$z$ 为$\left\{\begin{matrix}\text{极大值}\\\text{极小值}\end{matrix}\right\}$

$$
\text{如果} \quad \begin{cases} |H_1|<0;\ |H_2|>0;\ |H_3|<0\ (d^2z\ \text{负定}) \\ |H_1|>0;\ |H_2|>0;\ |H_3|>0\ (d^2z\ \text{正定}) \end{cases}
\tag{11.19}
$$

在运用此条件时,我们必须在稳定点 $f_1=f_2=f_3=0$ 处,计算全部主子式的值。

当然,我们也可以应用特征根检验,把 $d^2z$ 的正定(负定)与334

海塞矩阵 $\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$ 所有特征根的正(负)联系起来。事实上,不将二阶充分条件描述为"二阶全微分 $d^2z$ 为正定(负定)",代之以描述"海塞矩阵 $H$(有别于海塞行列式 $|H|$)为正定(负定)",也是可以的。但是,在这种用法中,要注意 $H$ 的有定符号是指与 $H$ 相联系的二次型 $d^2z$ 的符号,而不是 $H$ 本身的元素的符号。

**例 1** 求函数

$$z = 2x_1^2 + x_1x_2 + 4x_2^2 + x_1x_3 + x_3^2 + 2$$

的极值。极值的一阶必要条件要求同时满足下面三个方程:

$$(f_1 =)4x_1 + x_2 + x_3 = 0$$

$$(f_2 =)x_1 + 8x_2 \qquad = 0$$

$$(f_3 =)x_1 \qquad + 2x_3 = 0$$

因为这是一个齐次线性方程组,其中三个方程均独立无关(系数矩阵行列式不为零),所以只有一个唯一解 $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 0$。这意味着仅有一个稳定值 $z = 2$。

此函数的海塞行列式为

$$|H| = \begin{vmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{vmatrix} = \begin{vmatrix} 4 & 1 & 1 \\ 1 & 8 & 0 \\ 1 & 0 & 2 \end{vmatrix}$$

其主子式均为正:

$$|H_1| = 4 \qquad |H_2| = 31 \qquad |H_3| = 54$$

因此由(11.9),我们可以得出结论:$\bar{z}=2$ 是极小值。

**例 2** 求函数

$$z=-x_1^3+3x_1x_3+2x_2-x_2^2-3x_3^2$$

的极值。求得其一阶偏导数如下:

$$f_1=-3x_1^2+3x_3 \quad f_2=2-2x_2 \quad f_3=3x_1-6x_3$$

令所有的 $f_i=0$,我们得到三个联立方程。其中一个是非线性的,
335 两个是线性的:

$$\begin{aligned} -3x_1^2 \qquad\qquad +3x_3 &= 0 \\ -2x_2 \qquad\qquad &= -2 \\ 3x_1 \qquad\qquad -6x_3 &= 0 \end{aligned}$$

因为第二个方程给出 $\bar{x}_2=1$,第三个方程意味着 $\bar{x}_1=2x_3$,将其代入第一个方程,得到两个解:

$$(\bar{x}_1,\bar{x}_2,\bar{x}_3)=\begin{cases}(0,1,0), \text{ 意味着 } \bar{z}=1 \\ (\frac{1}{2},1,\frac{1}{4}), \text{ 意味着 } \bar{z}=\frac{17}{16}\end{cases}$$

适当排列二阶偏导数,得到海塞行列式

$$|H|=\begin{vmatrix} -6x_1 & 0 & 3 \\ 0 & -2 & 0 \\ 3 & 0 & -6 \end{vmatrix}$$

其中,根据第一个解($\bar{x}_1=0$),第一个元素($-6x_1$)可简化为零;根据第二个解($\bar{x}=1/2$)此元素可简化为 $-3$。立即可以看出,第一个解不满足二阶充分条件,因为 $|H_1|=0$。但我们可以借助于特征根检验对此进行进一步验证。为此,我们应用特征方程(11.14)。因为用于检验的二次型是 $d^2z$,其判别式是海塞行列式,当然,我们应当用海塞行列式中的元素代替方程中的元素 $d_{ij}$。因

此，第一个解的特征方程为

$$\begin{vmatrix} -r & 0 & 3 \\ 0 & -2-r & 0 \\ 3 & 0 & -6-r \end{vmatrix} = 0$$

将其展开，变成一个三次方程

$$r^3 + 8r^2 + 3r - 18 = 0$$

用试错法，我们将三次函数因式分解，将上面的方程重写成

$$(r+2)(r^2+6r-9) = 0$$

显然，由$(r+2)$项可知，一个特征根为 $r_1 = -2$。对另一项运用平方公式，可求出另外两个特征根：$r_2 = -3 + \frac{1}{2}\sqrt{72}$，$r_3 = -3 - \frac{1}{2}\sqrt{72}$。由于 $r_1, r_3$ 为负，而 $r_2$ 为正，二次型 $d^2z$ 是不定的，因此违背 $z$ 的极大值和极小值的二阶必要条件。因此，第一个解($\bar{z} = 1$)不是极值，只是一个拐点。

至于第二个解，情况便简单了。因为逐次主子式为

$$|H_1| = -3 \qquad |H_2| = 6 \qquad 和 \ |H_3| = -18$$

符号正好交替改变，行列式检验能够得出结论。根据(11.19)，解 336 $\bar{z} = 17/16$ 是极大值。

## n－变量的情况

当存在 $n$ 个变量时，目标函数可以表示成

$$z = f(x_1, x_2, \cdots, x_n)$$

全微分则为

$$dz = f_1 dx_1 + f_2 dx_2 + \cdots + f_n dx_n$$

所以极值的必要条件(对任意 $dx_i, dz=0$)意味着要求所有的一阶偏导数等于零。

二阶微分 $d^2z$ 还是一个二次型,其推导过程类似于(11.18),并可用一个 $n\times n$ 阵列来表示。适当重排该阵列的系数,可得到对称海塞行列式:

$$|H| = \begin{vmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{vmatrix}$$

其主子式 $|H_1|, |H_2|, \cdots, |H_n|$ 的定义如前。同以前一样,极值的二阶充分条件是:对于 $z$ 的极小值,所有 $n$ 个主子式为正;对于 $z$ 的极大值,第一个主子式为负,其它主子式符号交替改变。

总而言之,若我们集中于行列式检验,则我们把检验标准列于表 11.2 中,对于有任意选择变量的目标函数,这些标准均是有效的。在特殊情况下,我们可能有 $n=1$ 或 $n=2$。当 $n=1$ 时,目标函数为 $z=f(x)$,最大化条件 $f_1=0$, $|H_1|<0$ 可简化为 $f'(x)=0, f''(x)<0$,同我们在 9.4 节中所学的内容完全一致。类似地,当 $n=2$ 时,目标函数为 $z=f(x_1, x_2)$,因而极大值的一阶条件为 $f_1=f_2=0$,而二阶充分条件变成

337

$$f_{11}<0 \quad \text{和} \quad \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} = f_{11}f_{22}-f_{12}^2>0$$

这不过是表 11.1 所列信息的重述而已。

表 11.2　**相对极值的行列式检验：$z=f(x_1, x_2, \cdots, x_n)$**

| 条　件 | 极大值 | 极小值 |
|---|---|---|
| 一阶必要条件 | $f_1=f_2=\cdots=f_n=0$ | $f_1=f_2=\cdots=f_n=0$ |
| 二阶充分条件① | $\lvert H_1\rvert<0$; $\lvert H_2\rvert>0$; $\lvert H_3\rvert<0$; $\cdots$; $(-1)^n\lvert H_n\rvert>0$ | $\lvert H_1\rvert, \lvert H_2\rvert, \cdots, \lvert H_n\rvert>0$ |

## 练习 11.4

求下列五个函数的极值(若存在的话)。运用行列式检验判定它们是极大值,还是极小值。

1　$z=x_1^2+3x_2^2-3x_1x_2+4x_2x_3+6x_3^2$

2　$z=29-(x_1^2+x_2^2+x_3^2)$

3　$z=x_1x_3+x_1^2-x_2+x_2x_3+x_2^2+3x_3^2$

4　$z=e^x+e^y+e^{w^2}-2e^w-(x+y)$

5　$z=e^{2x}+e^{-y}+e^{w^2}-(2x+2e^w-y)$

然后回答下列关于其海塞矩阵和其特征根的问题：

6　($a$) 上述五题哪些产生了对角海塞矩阵？在每个对角海塞矩阵中，对角元素拥有相同符号吗？

　($b$) 对求得的每个对角海塞阵,关于其特征根,你能得出什么结论？

　($c$) 特征根检验的结果与行列式检验的结果相符吗？

7　($a$) 求第 3 题海塞矩阵的特征根。

　($b$) 你对求得的结果能得出什么结论？

　($c$) 你对于($b$)问题的答案是否与前面用行列式检验法对第 3 题的检验结论相一致？

---

① 仅当一阶必要条件得到满足后,方能应用此条件。

## 11.5 与函数凹性和凸性相关的二阶条件

无论是按照海塞行列式主子式表述的二阶条件，还是按照海塞矩阵特征根表述的二阶条件，总是与稳定点是峰顶还是谷底这一问题有关。换言之，它们总是与一条曲线、一个曲面或者一个超曲面(视情况而定)在稳定点附近如何弯曲有关。在单一选择变量
338 情况下，即 $z=f(x)$ 情况下，峰顶或谷底的图形是以一条倒 $U$ 形或 $U$ 形曲线表示的。对于二元函数 $z=f(x,y)$，其峰顶(谷底)的形状是以山丘形(碗形)表面来表示的，如图 11.2$a$(图 11.2$b$)所示。当存在三个或更多选择变量时，峰顶或谷底难以图形化，但我们可以在超平面中想象出“峰顶”或“谷底”。

一个在整个定义域中给出峰形(谷底)的函数被称作凹(凸)函数。① 在现在的讨论中，我们取定义域为整个 $R^n$，其中 $n$ 为选择变量数。由于峰形或谷底在整个定义域中存在，所以凹性和凸性当然也就是一个整体概念。为便于更好的分类，我们一方面在凹性和凸性之间作了区分，另一方面也在**严格**凹性和**严格**凸性之间作了区分。在**非严格**的情况下，允许峰形或谷底包含一个或多个平坦(相对于弯曲)的部分，比如线段(在曲线上)或者线段和平面(在曲面上)。但“严格”一词的存在，则排除了线段或平面存在的可能性。图 11.2 所示的两个曲面，分别代表严格凹函数和严格凸函数。而在图 6.5 中的曲线，是一条凸曲线，但不是严格凸的曲线(它包含了一段线段)。严格凹(凸)函数必定为凹函数，但反之不

① 若山峰(谷底)仅在定义域的子集 $S$ 中出现，则称函数在子集 $S$ 上凹(凸)。

成立。

考虑到凹性和严格凹性与整体峰形之间的联系,凹函数的极值必然是极大值——峰顶。而且,此极大值必定为绝对极大值(与相对极大值对照),因为峰形覆盖整个定义域。但绝对极大值可能不是唯一的,因为如果山峰包含一个平顶,则可能存在多重绝对极大值。仅当我们限定为严格凹性时,才可以排除后一种可能性。只有如此,峰值才包括一个单一的点,绝对极大值才是唯一的。唯一的(非唯一的)绝对极大值也称作强(弱)绝对极大值。

通过类似推理可知,凸函数的极值必定为绝对(或整体)极小值,但可能不是唯一的。但严格凸函数的极值必定是唯一的绝对极小值。

在上一段中,凹性和凸性的性质是在整体范围中考察的。若这些性质仅在曲线或曲面(仅在定义域中的子集 $S$)的一部分中成立,则与之相联系的极大值或极小值,则是相对于定义域该子集的极值,因为我们不能确定子集 $S$ 以外的情况。在我们前面关于 $d^2z$(或海塞矩阵 $H$)有定符号的讨论中,我们仅在稳定点计算海塞行列式主子式的值。因此,通过把山峰或谷底形状的验证限定在稳定点非常小的邻域内,我们仅能讨论相对极大和相对极小。但也可能恰好无论主子式在何处计值,$d^2z$ 都有确定符号。在此情况下,峰形或谷底将覆盖整个定义域,所求的极大值或极小值应为绝对极值。更具体地,若 $d^2z$ 处处为半负(正)定,则函数 $z=f(x_1, x_2, \cdots, x_n)$必定为凹(凸)函数;若 $d^2z$ 处处为负(正)定,则 $f$ 必为严格凹(严格凸)函数。 339

上面关于二次连续可微函数 $z=f(x_1, x_2, \cdots, x_n)$的讨论归纳在图 11.5 中。为明确起见,我们仅集中于凹性和极大值;但

如果“凹”、“负”、“极大值”分别被“凸”、“正”、“极小值”等词替代，所描述的关系依然成立。在阅读图 11.5 时，回忆一下，符号⇒(这

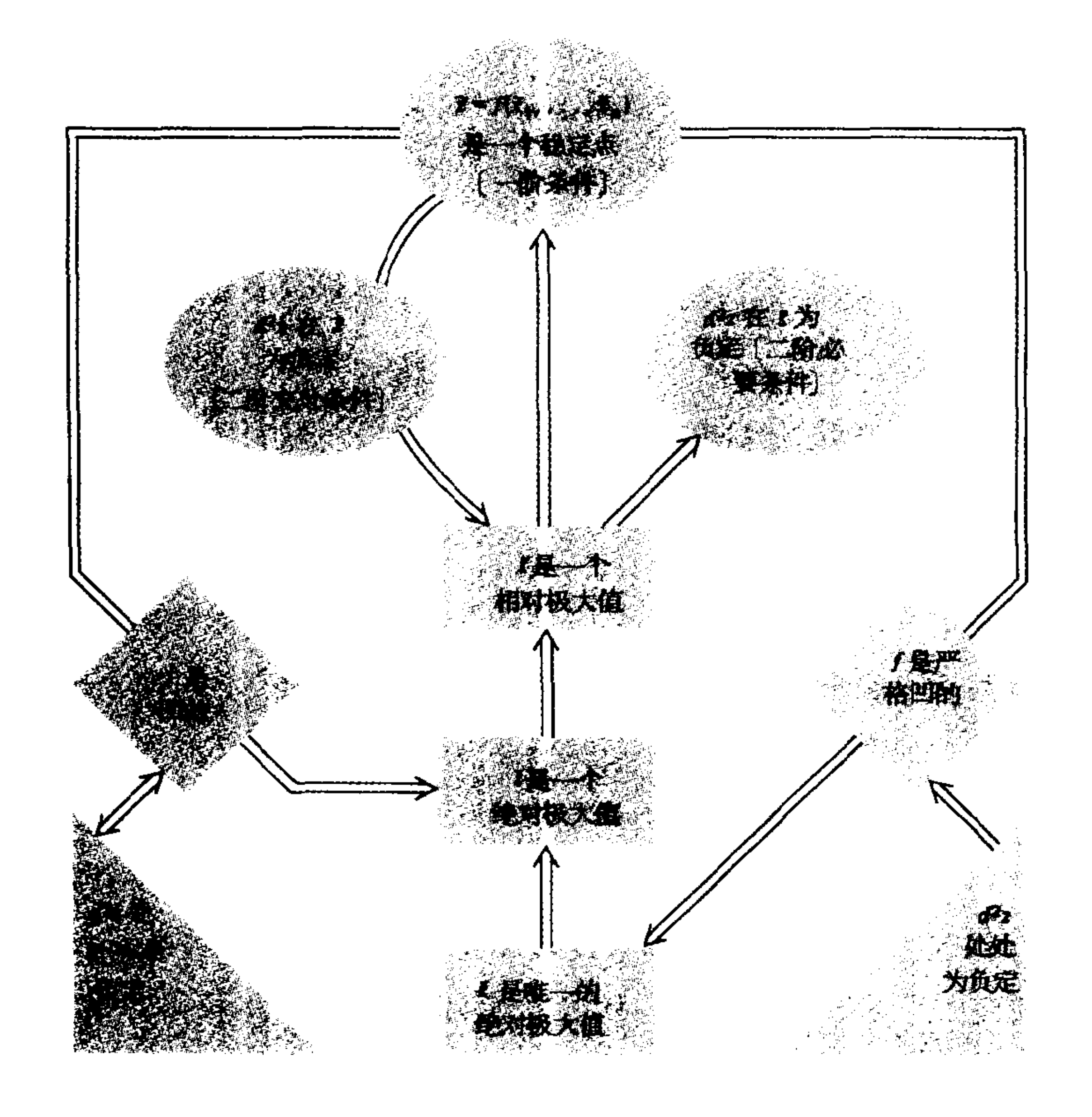



图　11.5

里符号被拉长、甚至弯曲)表示“意味着”。当该符号从一个单元(比如一个矩形)延伸至另一单元(比如椭圆形)时，它表示前者意味着后者(前者对后者是充分的)；它也表示后者对前者是必要的。当符号⇒从一单元通过第二个单元延至第三个单元时，它表示第

340 一个单元与第二个单元一起，意味着第三个单元。

据此，图 11.5 中间一列（从上向下读）表明：一阶条件是 $\bar{z}$ 为相对极大值的必要条件；而 $\bar{z}$ 为相对极大值又是 $\bar{z}$ 为绝对极大值的必要条件，等等。反之，将中间那列从下向上读，我们看到：$\bar{z}$ 为唯一的绝对极大值是确定 $\bar{z}$ 为绝对极大值的充分条件，进而又是 $\bar{z}$ 为相对极大值的充分条件，如此等等。最上边的三个椭圆形与在稳定点 $\bar{z}$ 的一阶和二阶条件有关。因此，它们仅与相对极大值相联系。下半部分的菱形和三角形，则描述了使我们可以得出关于绝对极大值结论的整体性质。注意，尽管我们前面的讨论表明 $d^2z$ 处处为半负定仅是函数 $f$ 为凹性的充分条件，但我们在图 11.5 中增加了这一信息：它也是 $f$ 为凹性的必要条件。与此相对照的是，$d^2z$ 处处为负定是函数 $f$ 为严格凹的充分条件，但不是必要条件，因为 $f$ 的严格凹性与 $d^2z$ 在稳定点取零值是一致的。

然而，穿过两个菱形的延长的⇒符号传递的信息是图 11.5 中最重要的信息。左边的一个表明：给定凹目标函数，任意稳定点可立即被视为一个绝对极大值。而右边的一个则表明：若目标函数是严格凹函数，稳定点则必定是唯一的绝对极大值。无论在哪种情况下，一旦一阶条件得到满足，凹性或严格凹性实际上取代二阶条件，成为极值（而是绝对极值）的充分条件。如果我们回顾一下，$d^2z$ 的峰值可能恰好为零，因而导致二阶充分条件失效，那么，我们就会更清楚地理解这一新的充分条件的有效性。凹性或严格凹性条件甚至可以处理这些有些麻烦的峰值，因为即使二阶条件未能得到满足，凹性或严格凹性能保证二阶以上的充分条件得到满足。正是由于这个原因，当我们构建一个具有一般目标函数的最优化模型时，我们从一开始就假定目标函数具有凹性（类似地，我们假定极小化模型的目标函数具有凸性）。但要注意，若使用具体

目标函数,则其凹凸性就不能作简单的假设,必须进行检验。

## 目标函数凹、凸性的检验

函数的凹性和凸性(严格的或不严格的)可用多种方式进行定义和检验。我们首先按照与 9.3 节类似的方式(对一元函数的情 341 况),定义双变量函数 $z=f(x_1, x_2)$的凹凸性:

> 对于函数 $z=f(x_1, x_2)$,在函数曲面上找任意两个不同的点 $M$ 和 $N$,连接线段 $MN$,当且仅当 $MN$ 位于曲面表面或曲面下方(上方)时,函数 $z$ 为凹(凸)函数。当且仅当除了点 $M$ 和 $N$ 外,线段 $MN$ 完全位于曲面下方(上方)时,函数 $z$ 为严格凹(严格凸)函数。

图 11.6 描绘了严格凹函数的情况。其中 $M$ 与 $N$ 为曲面上

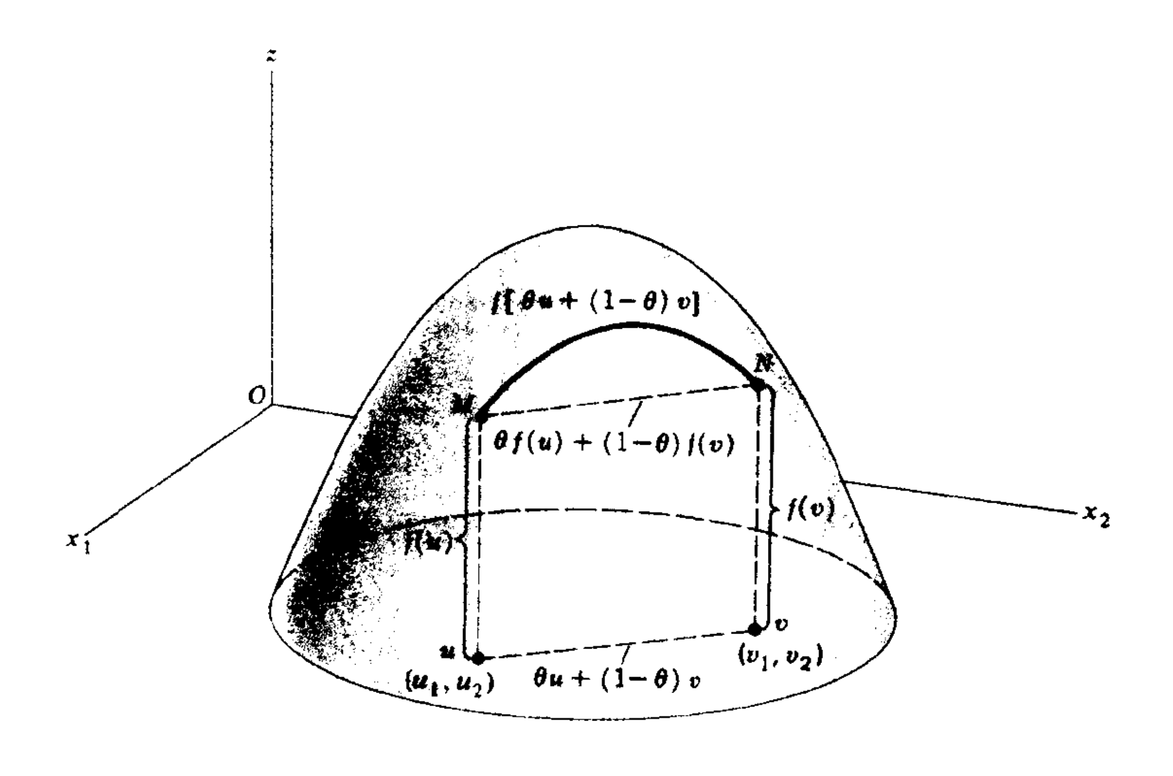



图 11.6

的两个任意点,由一条虚线段和一条实弧线连结起来,实弧线由

位于虚线段正上方的曲面上的点的集合组成。因为对于曲面上任意两点 $M$ 和 $N$,严格凹性要求线段 $MN$(除点 $M$ 和 $N$ 外)完全位于弧线 $MN$ 之下,所以曲面必定为山丘状的。类似地,严格凸函数的图形必定为碗形的。至于非严格凹和凸函数,因允许线段 $MN$ 位于曲面上(曲面的一部分,甚至整个曲面),其图形可能是一个平面,而非曲面。

为便于推广到不能图形化的 $n$ 维情况,我们需要将几何定义转换为等价的代数形式。回到图 11.6,令 $u=(u_1,\ u_2)$ 和 $v=(v_1,\ v_2)$ 为 $z=f(x_1,\ x_2)$ 定义域内的任意两个不同的有序偶(2-向量),则对应于它们的 $z$ 值(曲面的高度)分别为 $f(u)=f(u_1,\ u_2)$ 和 $f(v)=f(v_1,\ v_2)$。我们已假定,变量可取所有实数值,所以,若 $u$ 与 $v$ 在定义域内,则线段 $uv$ 上所有的点也都在 342 定义域内。而线段 $uv$ 上所有的点,实质上是 $u$ 和 $v$ 的"加权平均"。因此,我们可以 $\theta u+(1-\theta)v$ 来表示此线段,其中 $\theta$(希腊字母 theta)不同于 $u$ 和 $v$,是一个取值范围为 $0\leqslant\theta\leqslant1$ 的可变的标量。[①] 同理,表示 $f(u)$ 和 $f(v)$ 加权平均值的所有点的集合的线段 $MN$,可用 $\theta f(u)+(1-\theta)f(v)$ 表示,$\theta$ 的取值范围仍为 0 到 1。那么,曲面上的弧线段 $MN$ 又如何呢?因为弧线表示函数 $f$ 在线段 $uv$ 上不同点计算的值,所以可以将弧线 $MN$ 简单地写成 $f[\theta u+(1-\theta)v]$。利用这些表达式,我们现在可以给出如下代数定义:

---

① 对于 0 和 1 之间的任意具体 $\theta$ 值,加权平均表达式 $\theta u+(1-\theta)v$,特指两个变量 $u$ 和 $v$ 的凸组合。对此概念我们在本节后面还要详细解释,这里仅需注意这一点:当 $\theta=0$ 时,给定表达式简化为向量 $v$;类似地,当 $\theta=1$ 时,表达式简化为向量 $u$。而在 0 与 1 间的 $\theta$ 值,则给出了两个向量 $u$ 和 $v$ 的平均值。

对于函数 $f$ 定义域内任意两个不同的点 $u$ 和 $v$，且对于 $0<\theta<1$，当且仅当

$$\underbrace{\theta f(u)+(1-\theta)f(v)}_{\text{线段的高度}} \quad \begin{Bmatrix} \leqslant \\ \geqslant \end{Bmatrix} \quad \underbrace{f[\theta u+(1-\theta)v]}_{\text{弧的高度}} \tag{11.20}$$

时，$f$ 为 $\begin{cases} \text{凹函数} \\ \text{凸函数。} \end{cases}$

注意，为将两个端点 $M$ 和 $N$ 从高度比较中排除，我们将 $\theta$ 限定在开区间(0,1)中。

将弱不等号"≤"和"≥"分别变换成严格不等号"<"和">"，上述定义便适应于**严格**凹性和凸性的定义。代数定义的优势在于，它们可以应用于任意变量数的函数，因为定义中的向量 $u$ 和 $v$ 完全可以解释成 $n$ - 向量，而非 2 - 向量。

由(11.20)，下面三个关于凹性和凸性的定理可以非常容易地推导出来。这些定理将按函数 $f(x)$ 和 $g(x)$ 来表述，但 $x$ 可以解释成向量变量，所以这些定理对具有任意变量数的函数均适用。

**定理 I (线性函数)** 若 $f(x)$ 是一个线性函数，则此函数既可以是凹函数，也可以是凸函数，但不是严格凹或凸函数。

**定理 II (函数的正负与凹凸性)** 若 $f(x)$ 为凹函数，则 $-f(x)$ 为凸函数，反之亦然；类似地，若 $f(x)$ 为严格凹函数，则 $-f(x)$ 为严格凸函数，反之亦成立。

**定理 III (函数的和)** 若 $f(x)$ 与 $g(x)$ 均为凹(凸)函数，则 $f(x)+g(x)$ 也为凹(凸)函数；若 $f(x)$ 和 $g(x)$ 均为凹(凸)函数，且其中至少有一个为严格凹(严格凸)函数，则 $f(x)+g(x)$ 为严格凹(严格凸)函数。

343

定理 I 的结论基于这样一个事实:线性函数的图形是一条直线,或是一个平面,或是一个超平面,所以“线段 $MN$”与“弧线 $MN$”总是重合的。因而,(11.20)式中的两个弱不等号中的等号部分同时成立,使得此函数既是一个凹函数,也是一个凸函数。但因它舍弃了定义中的严格不等式部分,所以线性函数既非严格凹函数,也非严格凸函数。

定理 II 则以这样的事实为基础:凹性和凸性的定义仅在不等号的方向上有所区别。假定 $f(x)$ 为凹函数,则

$$\theta f(u)+(1-\theta)f(v)\leqslant f[\theta u+(1-\theta)v]$$

以 $-1$ 通乘,使不等号完全改变方向,则我们得到

$$\theta[-f(u)]+(1-\theta)[-f(v)]\geqslant -f[\theta u+(1-\theta)v]$$

而这恰好为 $-f(x)$ 为凸函数的条件。因此,对于凹函数 $f(x)$ 的情况,定理得到了证明。此结果的几何解释也是非常简单的:相对于底平面或超平面的山峰的镜像是深谷。其它情况也可通过类似办法证明。

为弄清定理 III 成立的理由,我们假定 $f(x)$ 和 $g(x)$ 均为凹函数。则下面两个不等式成立:

$$\theta f(u)+(1-\theta)f(v)\leqslant f[\theta u+(1-\theta)v] \tag{11.21}$$

$$\theta g(u)+(1-\theta)g(v)\leqslant g[\theta u+(1-\theta)v] \tag{11.22}$$

使其相加,得到新不等式

$$\begin{aligned}&\theta[f(u)+g(u)]+(1-\theta)[f(v)+g(v)]\\&\quad\leqslant f[\theta u+(1-\theta)v]+g[\theta u+(1-\theta)v]\end{aligned} \tag{11.23}$$

而此式恰好为 $[f(x)+g(x)]$ 为凹函数的条件。因此对于凹函数的情况,此定理得到了证明。凸函数情形的证明也类似。

现在考察定理 III 的第二部分内容，令 $f(x)$ 为严格凸函数。则(11.21)式变成严格不等式：

$$\theta f(u)+(1-\theta)f(v)<f[\theta u+(1-\theta)v] \quad (11.21')$$

将此式与(11.22)相加，我们看到两个不等式左边表达式之和严格小于右边表达式之和，无论(11.22)中的<号成立，还是=号成立，这意味着(11.23)现在变成了一个严格不等式，因而使得$[f(x)+g(x)]$严格凹。此外，若 $g(x)$与 $f(x)$一起成为严格凹函数，即若(11.22)与(11.21)一道转换成严格不等式，则可得出同样的结论。这样，对凹函数的情况，证明了定理的第二部分。凸函数情况的证明也类似。

344 对于两个以上凹(凸)函数的和，定理 III 依然成立。此定理有时很有价值，因为当一个函数包含相加项时，它使检验函数凹凸性的任务分解成为可能。如果已知函数相加的各项均为凹(凸)，则可知道和函数必然也为凹(凸)。

**例1** 检验 $z=x_1^2+x_2^2$ 的凹凸性。为应用(11.20)，令 $u=(u_1, u_2)$，$v=(v_1, v_2)$为定义域中任意两个不同的点。则我们有：

$$f(u)=f(u_1, u_2)=u_1^2+u_2^2$$

$$f(v)=f(v_1, v_2)=v_1^2+v_2^2$$

和 $$f[\theta u+(1-\theta)v]=f[\underbrace{\theta u_1+(1-\theta)v_1}_{x_1\text{的值}},\underbrace{\theta u_2+(1-\theta)v_2}_{x_2\text{的值}}]$$

$$=[\theta u_1+(1-\theta)v_1]^2+[\theta u_2+(1-\theta)v_2]^2$$

代入(11.20)，从左边表达式中减去右边表达式并合并各项，求得其差为

$$\theta(1-\theta)(u_1^2+u_2^2)+\theta(1-\theta)(v_1^2+v_2^2)$$
$$-2\theta(1-\theta)(u_1v_1+u_2v_2)$$
$$=\theta(1-\theta)[(u_1-v_1)^2+(u_2-v_2)^2]$$

因$\theta$是一个正分数，$\theta(1-\theta)$必然为正。进而因$(u_1, u_2)$和$(v_1, v_2)$为不同的点，所以或者有$u_1 \neq v_1$，或者有$u_2 \neq v_2$（或者二者都不相等），因而括号内的表达式必定也为正。因此，(11.20)中的严格>号成立，且$z=x_1^2+x_2^2$为严格凸函数。

另外，我们还可以单独检验$x_1^2$项和$x_2^2$项。因为这两项中每一项都是严格凸的，所以其和也是严格凸的。

因为此函数为严格凸函数，所以它具有唯一的绝对极小值。很容易验证，所说的最小值是$\bar{z}=0$，此最小值在$\bar{x}_1=\bar{x}_2=0$达到，而且它确实是唯一的绝对最小值，因为任意有序偶$(x_1, x_2) \neq (0, 0)$所产生的$z$值都大于零。

**例2**　检验$z=-x_1^2-x_2^2$的凹凸性。此函数为例1函数的负数。因此，根据定理II，它是严格凹函数。

**例3**　检验$z=(x+y)^2$的凹凸性。尽管这里变量以$x$和$y$表示而不以$x_1$、$x_2$表示，但我们仍可用$u=(u_1, u_2)$和$v=(v_1, v_2)$表示定义域中的两个不同的点，用下标$i$表示第$i$个变量。这样，我们有

345

$$f(u)=f(u_1, u_2)=(u_1+u_2)^2$$
$$f(v)=f(v_1, v_2)=(v_1+v_2)^2$$

和　$$f[\theta u+(1-\theta)v]=[\theta u_1+(1-\theta)v_1+\theta u_2+(1-\theta)v_2]^2$$
$$=[\theta(u_1+u_2)+(1-\theta)(v_1+v_2)]^2$$

将其代入(11.20)，从左边表达式中减去右边表达式，并简化，我们

求得其差为

$$\theta(1-\theta)(u_1+u_2)^2-2\theta(1-\theta)(u_1+u_2)(v_1+v_2)$$
$$+\theta(1-\theta)(v_1+v_2)^2$$
$$=\theta(1-\theta)[(u_1+u_2)-(v_1+v_2)]^2$$

同例 1 中一样,$\theta(1-\theta)$为正,括号中的表达式的平方为非负(这时不能排除其为零)。因此,(11.20)中的大于等于号≥成立,函数$(x+y)^2$ 为凸函数,但不是严格凸函数。

因此,此函数具有一个绝对极小值,此极小值在任意$(\bar{x}+\bar{y})=0$ 达到,它可能并不是唯一的。很明显,它是一个绝对极小值,因为只要 $x+y\neq 0$,$z$ 就将大于 $\bar{z}=0$。它不是唯一的极小值则是由于存在无数个可以满足条件 $\bar{x}+\bar{y}=0$ 的有序偶$(\bar{x},\bar{y})$。

## 可微函数

在(11.20)中所描述的函数凹凸性定义并未用到导数概念,因而并不要求函数具有可微性。但如果函数是可能的,函数的凹凸性也可以按其一阶导数来定义。在单变量情况下,其定义为:

对于定义域中任意两个不同的给定点 $u$ 和 $v$,当且仅当

$$f(v)\begin{Bmatrix}\leqslant\\ \geqslant\end{Bmatrix}f(u)+f'(u)(v-u) \qquad (11.24)$$

时,可微函数 $f(x)$ 为$\begin{cases}\text{凹函数}\\ \text{凸函数。}\end{cases}$

如果(11.24)中的弱不等号分别被严格不等号<与>所代替,上面的定义就变成严格凹性和凸性的定义。从几何上看,此定义将凹(凸)曲线描绘成一条与其切线重合或者位于其切线下面(上面)的曲线。而严格凹曲线(严格凸曲线)必须位于所有切线的下

方（上方），但切点除外。

在图 11.7 中，令 $A$ 为曲线上任意一点，其高度 $f(u)$，过该点的切线为 $AB$。令 $x$ 自 $u$ 值开始递增，则为形成一个峰形，严格凹曲线必须逐渐向下弯曲，远离切线 $AB$，从而使点 $C$，其高度为

346

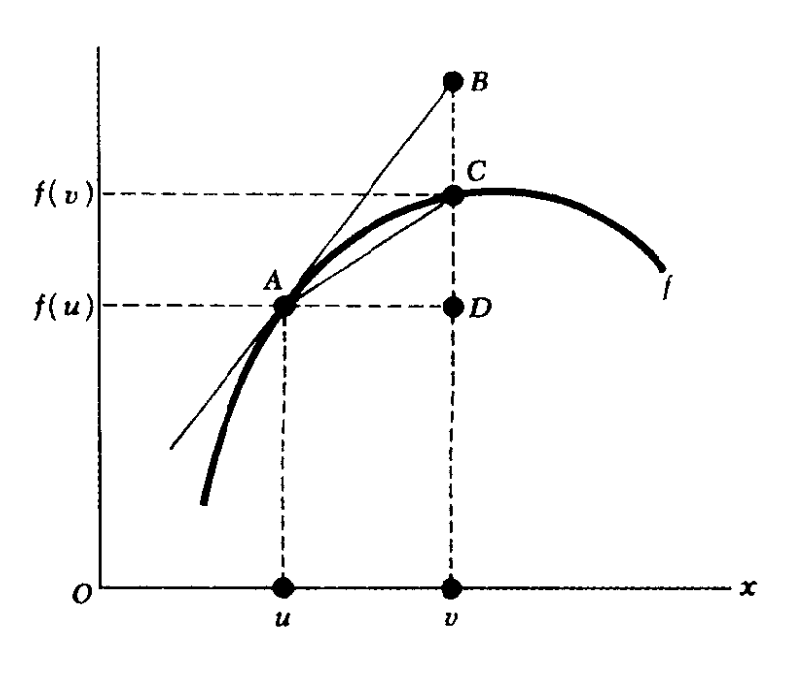



图 11.7

$f(v)$，必然位于点 $B$ 之下。在此情况下，线段 $AC$ 的斜率小于切线 $AB$ 的斜率。另一方面，若曲线为非严格凹曲线，它也可能包含一个线段，因而，比如弧线 $AC$ 可能变成一个线段，并与线段 $AB$ 重合，成为曲线中的一个直线部分。在后一种情况下，$AC$ 的斜率等于 $AB$ 的斜率。将这两种情况综合起来，它们意味着

$$(\text{线段 } AC \text{ 的斜率} = \frac{DC}{AD} = )\frac{f(v)-f(u)}{v-u} \leqslant (AB \text{ 的斜率} = ) = f'(u)$$

以正数 $(v-u)$ 通乘此不等式，得到一个在 (11.24) 中表达的适用于凹函数的结论。若我们考察小于 $u$ 的 $x$ 值，也可以得到同样的结论。

当函数中有两个或两个以上自变量时,此定义需要小小的修正:

对于定义域中任意给定点 $u=(u_1, \cdots, u_n)$和另一个给定点 $v=(v_1, \cdots, v_n)$,当且仅当

$$f(v) \begin{Bmatrix} \leqslant \\ \geqslant \end{Bmatrix} f(u)+\sum_{j=1}^{n} f_j(u)(v_j-u_j) \tag{11.24'}$$

可微函数 $f(x_1, \cdots, x_n)$为$\begin{cases}\text{凹函数}\\ \text{凸函数。}\end{cases}$ 其中 $f_j(u) \equiv \partial f / \partial x_j$ 在 $u=(u_1, \cdots, u_n)$计算其值。

此定义要求凹(凸)函数 $f(x)$的图形与其切平面或切超平面重合或位于其下方。对于严格凹性和严格凸性的情况,(11.24')中的弱不等式应换成严格不等式,它要求严格凹(严格凸)函数的图形严格位于所有切平面或切超平面之下(之上),但切点除外。

347

最后,考察一个二次连续可微的函数 $z=f(x_1, \cdots, x_n)$。因为此函数的二阶偏导数存在,因此 $d^2 z$ 有定义。所以函数凹性和凸性可用 $d^2 z$ 的符号来检验:

当且仅当 $d^2 z$ 处处为$\begin{Bmatrix}\text{负}\\ \text{正}\end{Bmatrix}$半定时,则二阶连续可微函数 $z=f(x_1, \cdots, x_n)$是$\begin{Bmatrix}\text{凹函数}\\ \text{凸函数}\end{Bmatrix}$。当(但不是仅当)$d^2 z$ 为是处处$\begin{Bmatrix}\text{负}\\ \text{正}\end{Bmatrix}$定时,所说的函数为严格$\begin{Bmatrix}\text{凹}\\ \text{凸}\end{Bmatrix}$函数。 (11.25)

读者应回想起来,(11.25)式关于凹函数和严格凹函数的有关内容已被纳入图 11.5 中。

**例 4** 运用导数条件检验函数 $z=-x^4$ 的凹凸性。我们首先应用(11.24)。在本例中,(11.24)左右两边的表达式分别为 $-v^2$ 和 $-u^4-4u^3(v-u)$。从前式中减去后式,我们得到其差为

$$-v^4+u^4+4u^3(v-u) \qquad [\text{提出公因子}]$$

$$=(v-u)(-\frac{v^4-u^4}{v-u}+4u^3)$$

$$=(v-u)[-(v^3+v^2u+vu^2+u^3)+4u^3] \quad [\text{由}(7.2)]$$

若括号内的表达式可被$(v-u)$除尽，那就美妙至极，因为这样我们就可以销掉$(v-u)$，并得到一个便于计算其符号的平方项$(v-u)^2$。正如将要证明的那样，上式确实如此。因此，上面所列的差可以写成

$$-(v-u)^2[v^2+2vu+3u^2]=-(v-u)^2[(v+u)^2+2u^2]$$

给定 $v\neq u$，此式的符号必定为负。(11.24)中的严格不等式成立，函数 $z=-x^4$ 是严格凹函数。这意味着它有唯一的绝对极大值。很容易验证，极大值为 $\bar{z}=0$，在点 $\bar{x}=0$ 处达到。

因为此函数是二阶连续可微的，所以我们也可以应用(11.25)来检验其凹凸性。由于仅有一个变量，因而(11.25)给出

$$d^2z=f''(x)dx^2=-12x^2dx^2 \quad [\text{由}(11.2)]$$

我们知道 $dx^2$ 为正(仅考察 $x$ 的非零变化)，但 $-12x^2$ 既可能为负，也可能为零。因此我们充其量得出这一结论：$d^2z$ 为处处负半定，且 $z=-x^4$ 为(非严格)凹函数。由(11.25)得出的结论显然弱于我们前面由(11.24)所得到的 $z=-x^4$ 是严格凹函数的结论。在这里使我们产生局限性，因而得出较弱结论的原因与导致二阶导数检验有时失效的原因是相同的，即 $d^2z$ 在已知严格凹函数或严格凸函数的稳定点可能取零值。这就是为什么在(11.25)中给出的 $d^2z$ 为负(正)半定仅是严格凹性(严格凸性)的充分条件，但不是必要条件的原因。 348

**例5** 以导数条件检验 $z=x_1^2+x_2^2$ 的凹凸性。这次我们用

(11.24′)而不用(11.24)。在定义域中有任意两点 $u=(u_1, u_2)$, $v=(v_1, v_2)$, (11.24′)的两边为

$$\text{左边} = v_1^2 + v_2^2$$

$$\text{右边} = u_1^2 + u_2^2 + 2u_1(v_1 - u_1) + 2u_2(v_2 - u_2)$$

从前者中减去后者,化简,我们可将差表示成:

$$v_1^2 - 2v_1u_1 + u_1^2 + v_2^2 - 2v_2u_2 + u_2^2 = (v_1 - u_1)^2 + (v_2 - u_2)^2$$

给定$(u_1, u_2) \neq (v_1, v_2)$,此差总为正。因此(11.24′)中的严格不等号成立,$z = x_1^2 + x_2^2$ 是严格凸函数。注意,这里的结论仅是再一次证实了我们前面在例 1 中已经得出了的结论。

若运用(11.25),因 $f_1 = 2x_1, f_2 = 2x_2$,无论二阶偏导数在何处计值,我们有

$$f_{11} = 2 > 0 \quad \text{和} \quad \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 > 0$$

因此,$d^2z$ 处处为正定,完全满足严格凸性的充分条件。因而在本例中,(11.24′)与(11.25)确实得出同样结论。

## 凸函数与凸集

在澄清了将形容词“凸的”一词应用于函数的含义之后,我们迫切需要解释用其描述集合的含义。尽管凸集与凸函数并非不相关,但它们却是不同的概念,所以,注意其区别,仅防混淆是重要的。

为便于直观把握,我们首先介绍凸集的几何特征。令 $S$ 为 2 维空间或 3 维空间中的点集,对于 $S$ 中的任意两点,若连接两点的线段完全位于 $S$ 内,则称 $S$ 为凸集。很明显,直线满足此条件

并构成了一个凸集。习惯上,仅包含一个点的集合也称为凸集,零集(没有点)也视为凸集。对于其它的例子,请考察图 11.8,圆盘 349

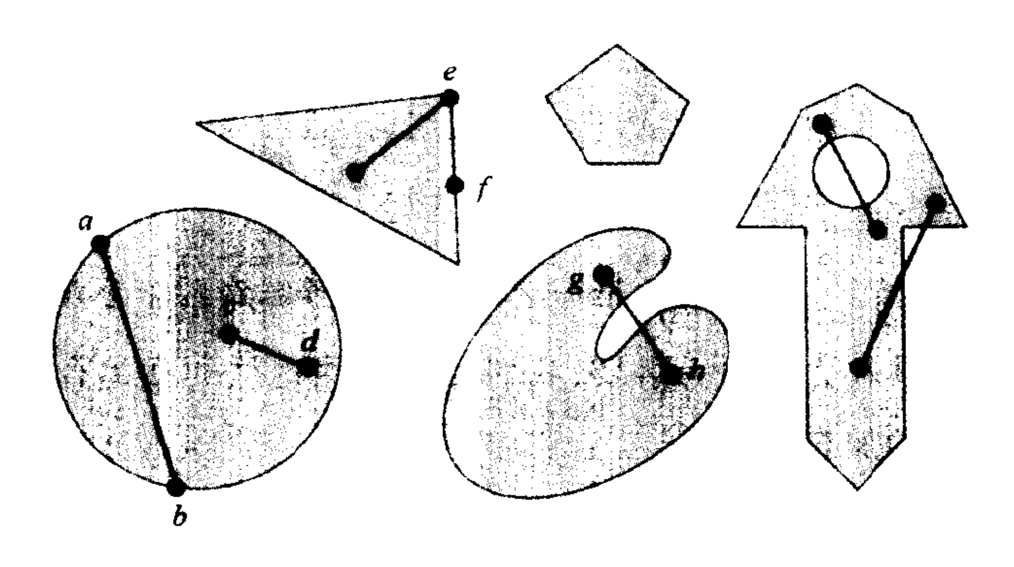



图 11.8

——即"实"圆,亦即一个圆加上其内部的所有点——是一个凸集,因为圆盘任意两点的连线均完全在圆盘内,比如图中的 $ab$(连接边缘上的两点)和 $cd$(连接圆内的两点)。但要注意,中空的圆本身不是凸集。类似地,三角形或五边形本身不是凸集,实的三角形或五边形则是凸集。在图 11.8 中余下的两个实的几何图形不是凸集。其中那个调色板型的图有一处凹入,因而像 $gh$ 这样的线段并不完全位于集合中。在那个钥形的图中,不仅有凹入,而且存在一个孔,这个孔又造成了另一种非凸性。一般而言,一个点集要成为凸集,此点集必须无孔,边缘各处无缩进。

凸性的几何定义也可以应用于 3 维空间中的点集。例如,一个实的立方体是凸集,而圆筒则不是凸集。但在四维或更高维空间中,几何解释变得不明显了。因而我们需要转向凸集的代数定义。

为此,引入**向量(点)凸组合**的概念是有益的。向量凸组合是

一种特殊类型的线性组合。两个向量 $u$ 和 $v$ 的凸组合可以写成:

$$k_1 u + k_2 v$$

其中 $k_1$ 和 $k_2$ 是两个标量。当两个标量均位于闭区间[0,1]内,且二者之和为 1 时,此线性组合称作凸组合,且可表示成

$$\theta u + (1-\theta)v \qquad (0 \leqslant \theta \leqslant 1) \tag{11.26}$$

举例来说,$\frac{1}{3}\begin{bmatrix}2\\0\end{bmatrix}+\frac{2}{3}\begin{bmatrix}4\\9\end{bmatrix}$是一个凸组合。考虑到两个标量乘子为正分数且其和为 1 这一事实,我们也可以将这样的凸组合解释成

350 两个向量的**加权平均**。①

组合(11.26)唯一的特征是,对于任意可接受的 $\theta$ 值,所产生的和向量位于连接点 $u$ 和 $v$ 的线段上。这一点可利用图 11.9 来

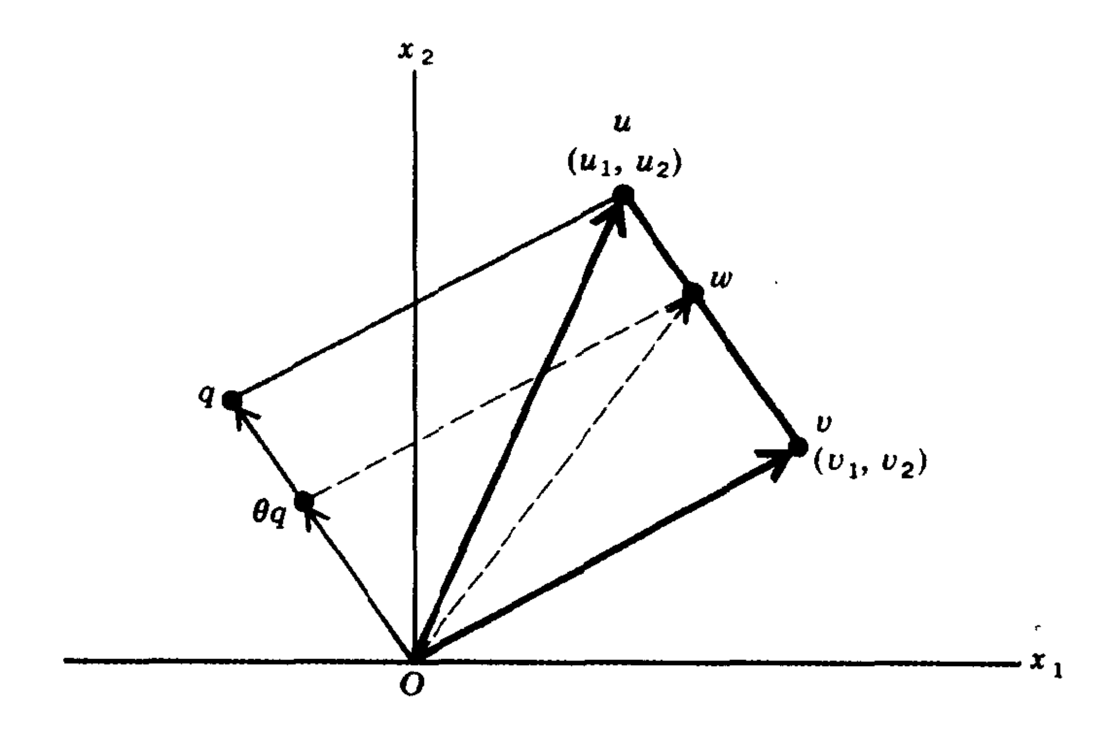



图 11.9

① 在讨论凹函数和凸函数时,我们已利用了这种解释方式。

证明。在图中我们把两个向量 $u=\begin{bmatrix}u_1\\u_2\end{bmatrix}$与 $v=\begin{bmatrix}v_1\\v_2\end{bmatrix}$绘成了分别与$(u_1,\ u_2)$和$(v_1,\ v_2)$相对应的两个点。若我们再绘出另一个向量 $q$,使得 $oquv$ 形成一个平行四边形,则根据在图 4.3 中的讨论,我们有

$$u=q+v \quad \text{或} \quad q=u-v$$

由此可知,向量 $u$ 和 $v$ 的凸组合(称其为 $w$)可用向量 $q$ 来表示,因为

$$w=\theta u+(1-\theta)v=\theta u+v-\theta v=\theta(u-v)+v=\theta q+v$$

因此,要绘出向量 $w$,我们只需用熟悉的平行四边形法相加 $\theta q$ 和 $v$ 即可。若标量 $\theta$ 为一正分数,向量 $\theta q$ 将仅是向量 $q$ 的简缩形式,因此 $\theta q$ 必然位于线段 $Oq$ 之上。所以使 $\theta q$ 和 $v$ 相加,我们必然会得到位于线段 $uv$ 上的线段 $w$,因为新的小一些的平行四边形只不过是原平行四边形的边 $qu$ 向下移而形成的。当然,向量 $w$ 的确切位置将随 $\theta$ 值的变化而变化。当 $\theta$ 由 0 变至 1,$w$ 的位置将由 $v$ 移至 $u$。因此,线段 $uv$ 上的所有点的集合,包括 $u$ 和 $v$ 本身,对应于向量 $u$ 和 $v$ 所有凸组合的集合。

基于上述分析,现在可将凸集定义如下:对于任意两点 $u\in S$ 351 和 $v\in S$,且对于每一个标量 $\theta\in[0,1]$,当且仅当 $w=\theta u+(1-\theta)v\in S$ 为真时,集合 $S$ 为凸集。因为此定义是代数形式的,所以不管向量 $u$ 和 $v$ 所在的空间是几维的,此定义均适用。将凸集与凸函数定义(11.20)相比较,我们看到尽管两个定义都使用同一形容词"凸的",但这个词在两个定义中却有根本不同的含义。在描述函数时,"凸的"一词确定一条曲线或曲面如何弯曲(它必然形成一个深谷)。但在描述一个集合时,此词确定集合中的点是如何"填

充”到一起的——不允许出现任何孔，边缘不能有缩进。因此凸集与凸函数有明显不同的数学内涵。

然而，凸函数与凸集并非毫不相关。首先，在定义凸函数时，我们需要定义域为凸集。这是因为定义(11.20)要求，对于定义域中的任意两点 $u$ 和 $v$，$u$ 和 $v$ 的所有凸组合——具体地 $\theta u+(1-\theta)v$，$0\leqslant\theta\leqslant 1$——必须也在定义域中，这实际上就是定义域必须为凸集的另一种表达方式。为满足这一要求，我们前面采纳了一个非常强的假设，即定义域由整个 $n$ - 空间构成(其中 $n$ 为选择变量数)，$n$ - 空间确实是一个凸集。但是掌握了凸集概念以后，我们完全可以弱化这一假设。因为我们需要假设定义域是 $R^n$ 空间中的一个凸子集，而非 $R^n$ 本身。

凸函数与凸集还存在另一种联系方式。若 $f(x)$ 为凸函数，则对任意常数 $k$，它可以引致一个凸集

$$S^{\leqslant}\equiv\{x \mid \mathrm{f}(x)\leqslant k\} \quad [f(x)] \tag{11.27}$$

图 11.10$a$ 描述了单变量的情形。集合 $S^{\leqslant}$ 由在虚水平线上或位于其下方的 $f(x)$ 曲线的弧段相联系的 $x$ 值组成。因此，它是横轴上以重黑点标示的线段，它是一个凸集。注意，如果 $k$ 值变化，集合 $S^{\leqslant}$ 会变成横轴上的不同线段，但它仍然是个凸集。

进一步，我们甚至还可以看到凹函数也以类似的方式与凸集相联系。首先，在(11.20)中凹函数之定义(像凸函数情况一样)是以定义域为凸集为基础的。进而，给定某一常数 $k$，即便凹函数，比如 $g(x)$，也可以产生一个相联系的凸集。此凸集为

$$S^{\geqslant}\equiv\{x \mid \mathrm{g}(x)\geqslant k\} \quad [g(x)\text{为凹函数}] \tag{11.28}$$

其中符号$\geqslant$代替了(11.27)中的符号$\leqslant$。从几何上看，对于单变量函数情形，如图 11.10$b$ 所示，集合 $S^{\geqslant}$ 包含了对应于与虚水平线

交叉或位于其上面的 $g(x)$ 曲线弧段的所有的 $x$ 值。因此,它仍是横轴上的一个线段——一个凸集。

尽管图 11.10 具体地描述了单变量的情形,在(11.27)和 352

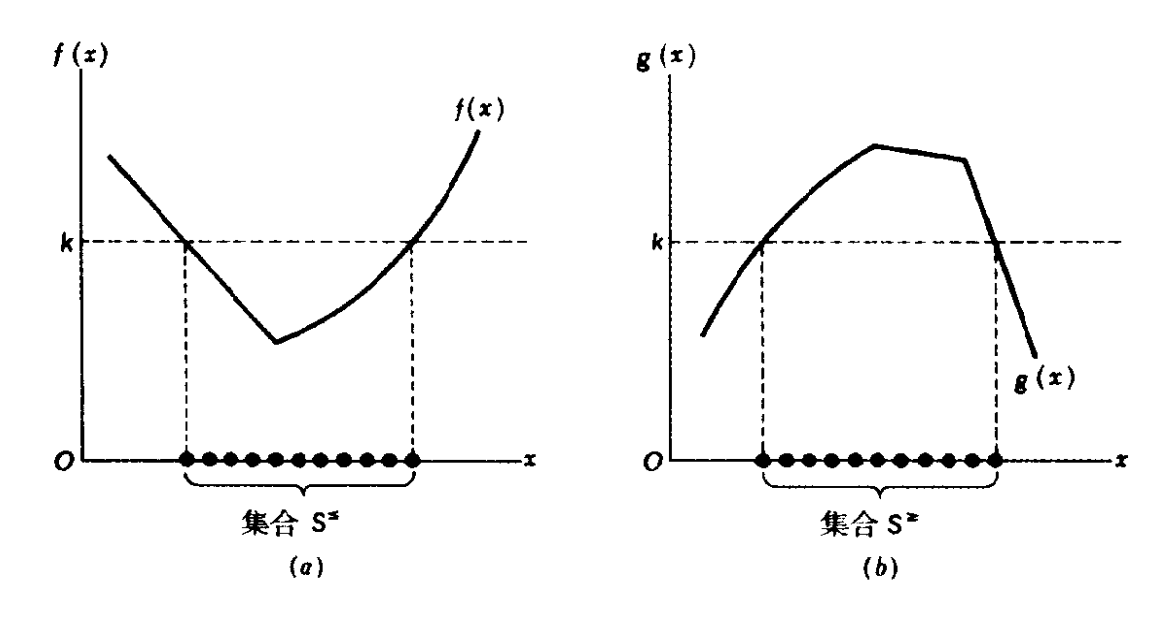



图 11.10

(11.28)中的 $S^{\leqslant}$ 与 $S^{\geqslant}$ 的定义却并不限于单变量函数。若我们将 $x$ 视作向量,即令 $x=(x_1, \cdots, x_n)$,它们同样成立。但在这种情况下,(11.27)和(11.28)将在 $n$-空间中定义凸函数。记住这一点是重要的:虽然凸函数意味着(11.27),凹函数意味着(11.28),但反之不成立,因为非凸函数也可满足(11.27),非凹函数也满足(11.28)。在 12.4 节中还要讨论这一问题。

## 练习 11.5

---

1 运用(11.20)检验下列函数是凹函数、凸函数,还是严格凹函数、严格凸函数,或者都不是:

$(a)z=x^2$ $(b)z=x_1^2+2x_2^2$ $(c)z=2x^2-xy+y^2$

2　运用(11.24)或(11.24′)检验下列函数是凹函数、凸函数,还是严格凹函数、严格凸函数,或者都不是:

$(a)z=-x^2$　$(b)z=(x_1+x_2)^2$　$(c)z=-xy$

3　根据你关于上题 $2c$ 的答案,你能运用定理 III 分解检验题 $1c$ 的函数 $z=2x^2-xy+y^2$ 这一任务吗? 解释你的结论。

4　在下列物体中,哪一种构成 3-空间中的凸函数?

(a) 汽车轮胎　(b) 保龄球　(c) 完好的大理石

5　方程 $x^2+y^2=4$ 表示以(0,0)为圆心,2 为半径的圆:

(a) 给出集合 $\{(x,y)/x^2+y^2\leqslant 4\}$ 的几何解释。

(b) 它是凸集吗?

353　6　绘出下列集合之图形,指出哪个是凸集:

$(a)\{(x,y)\mid y=e^x\}$　　$(c)\{(x,y)\mid y\leqslant 13-x^2\}$

$(b)\{(x,y)\mid y\geqslant e^x\}$　　$(d)\{(x,y)\mid xy\geqslant 1; x>0, y>0\}$

7　已知 $u=\begin{bmatrix}10\\6\end{bmatrix}$ 和 $v=\begin{bmatrix}4\\8\end{bmatrix}$,下列各式哪个是 $u$ 与 $v$ 的凸组合?

$(a)\begin{bmatrix}7\\7\end{bmatrix}$　$(b)\begin{bmatrix}5.2\\7.6\end{bmatrix}$　$(c)\begin{bmatrix}6.2\\8.2\end{bmatrix}$

8　已知 2-空间中的两个向量 $u$ 和 $v$,求出下列集合并绘出草图:

(a) $u$ 与 $v$ 所有线性组合的集合

(b) $u$ 与 $v$ 所有非线性组合的集合

(c) $u$ 与 $v$ 所有凸组合的集合

9　(a)按 $f$ 与 $g$ 函数具有 $n$ 个自变量的具体情况重写(11.27)和(11.28)。

(b)令 $n=2$,令函数 $f$ 的形状像一个竖握的锥形蛋卷冰激凌,函数 $g$ 的形状类似于一个棱锥体。描述集合 $S^{\leqslant}$ 与 $S^{\geqslant}$。

---

## 11.6　经济应用

本章开头,我们以多产品产商为例描述了具有多于一个选择变量的一般最优化问题。现在我们就准备处理这些问题及与其性

质类似的问题。

## 多产品厂商问题

**例 1**　我们首先假设有一个完全竞争条件下的两产品厂商。因完全竞争条件下两商品的价格必然是外生的，所以它们可分别以 $P_{10}$ 与 $P_{20}$ 表示。据此，厂商的收益函数为

$$R = P_{10}Q_1 + P_{20}Q_2$$

其中 $Q_i$ 表示单位时间内 $i$ 产品的产出水平。假设厂商成本函数为

$$C = 2Q_1^2 + Q_1Q_2 + 2Q_2^2$$

注意 $\partial C/\partial\theta_1 = 4Q_1 + Q_2$（第一个产品的边际成本）不仅是 $Q_1$ 的函数，而且是 $Q_2$ 的函数。类似地，第二个产品的边际成本也部分依赖于第一个产品的产出水平。因此，按照假定的成本函数，这两个商品在生产上看来存在技术的相关性。

现在可将此假定厂商的利润函数写成

$$\pi = R - C = P_{10}Q_1 + P_{20}Q_2 - 2Q_1^2 - Q_1Q_2 - 2Q_2^2$$

它是两个选择变量 $Q_1$、$Q_2$ 及两个价格参数的函数。我们的任务 354 是求出使$\pi$最大化的产出水平 $Q_1$ 与 $Q_2$ 的组合。为此，我们先求出利润函数的一阶偏导数

$$\begin{aligned} \pi_1 \ \left(\equiv \frac{\partial \pi}{\partial Q_1}\right) &= P_{10} - 4Q_1 - Q_2 \\ \pi_2 \ \left(\equiv \frac{\partial \pi}{\partial Q_2}\right) &= P_{20} - Q_1 - 4Q_2 \end{aligned} \qquad (11.29)$$

令二者等于零，为满足最大化的必要条件，我们得到联立方程

$$4Q_1 + Q_2 = P_{10}$$

$$Q_1 + 4Q_2 = P_{20}$$

产生唯一解

$$\bar{Q}_1 = \frac{4P_{10} - P_{20}}{15} \quad 和 \quad \bar{Q}_2 = \frac{4P_{20} - P_{10}}{15}$$

因此,若 $P_{10} = 12, P_{20} = 18$,我们有 $\bar{Q}_1 = 2, \bar{Q}_2 = 4$,这意味着单位时间的最大利润为 $\bar{\pi} = 48$。

为确认此值的确是最大利润,我们来检验二阶条件。对(11.29)偏导数得到的二阶偏导数给我们如下海塞行列式

$$|H| = \begin{vmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{vmatrix} = \begin{vmatrix} -4 & -1 \\ -1 & -4 \end{vmatrix}$$

因 $|H_1| = -4 < 0, |H_2| = 15 > 0$,海塞矩阵(或 $d^2z$)为负定,此解确实使利润最大化。事实上,主子式的符号与其在何处计值无关,所以在本例中,$d^2z$ 处处为负定。因此,根据(11.25),目标函数必定为严格凹函数,上面所求得的最大利润实际上是唯一的绝对极大值。

**例 2** 现在我们把例 1 移植到垄断市场环境中。由于这一新的市场结构假设,收益函数必须修正以反映这样的事实:两产品价格将随其产出水平(这里假设产出水平与销售水平一致,不考虑存货积累。)的变化而变化。当然,价格随产出水平变化的确切方式还有待于从对厂商两种产品的需求函数中求出。

假设对垄断厂商产品的需求函数如下:

$$\begin{aligned} Q_1 &= 40 - 2P_1 + P_2 \\ Q_2 &= 15 + P_1 - P_2 \end{aligned} \tag{11.30}$$

355 这两个方程揭示出,这两种商品在消费中存在着某种联系。具体地说,它们是替代品,因为一种商品价格的提高将提高对另一商品

的需求。正如(11.30)给出的那样,它表明需求量 $Q_1$ 和 $Q_2$ 是价格的函数,但就我们现在的目的而言,将价格 $P_1$ 和 $P_2$ 表示成 $Q_1$ 和 $Q_2$ 的函数,即两个产品的平均收益函数要更方便一些。因为(11.30)可以重写成

$$-2P_1 + P_2 = Q_1 - 40$$
$$P_1 - P_2 = Q_2 - 15$$

将 $Q_1$、$Q_2$ 视为参数,我们可应用克莱姆法则解 $P_1$ 和 $P_2$ 如下:

$$\begin{aligned} P_1 &= 55 - Q_1 - Q_2 \\ P_2 &= 70 - Q_1 - 2Q_2 \end{aligned} \qquad (11.30')$$

因为 $P_1 \equiv AR_1$,$P_2 \equiv AR_2$,所以这两个函数构成了所求的平均收益函数。

因而,厂商的总收益函数可以写成

$$\begin{aligned} R &= P_1Q_1 + P_2Q_2 \\ &= (55 - Q_1 - Q_2)Q_1 + (70 - Q_1 - 2Q_2)Q_2 \qquad [\text{由}(11.30')] \\ &= 55Q_1 + 70Q_2 - 2Q_1Q_2 - Q_1^2 - 2Q_2^2 \end{aligned}$$

若我们再假设总成本函数为

$$C = Q_1^2 + Q_1Q_2 + Q_2^2$$

则利润函数将为

$$\pi = R - C = 55Q_1 + 70Q_2 - 3Q_1Q_2 - 2Q_1^2 - 3Q_2^2 \qquad (11.31)$$

这是一个有两个选择变量的目标函数。一旦求出利润最大化的产出水平 $\bar{Q}_1$ 和 $\bar{Q}_2$,最优价格水平 $\bar{P}_1$ 和 $\bar{P}_2$ 便可由(11.30′)轻松求出。

目标函数产生如下两个一阶和二阶偏导数:

$$\pi_1 = 55 - 3Q_2 - 4Q_1 \qquad \pi_2 = 70 - 3Q_1 - 6Q_2$$

$$\pi_{11} = -4 \quad \pi_{12} = \pi_{21} = -3 \quad \pi_{22} = -6$$

为满足 $\pi$ 最大化的一阶条件，我们必须有$\pi_1 = \pi_2 = 0$，即

$$4Q_1 + 3Q_2 = 55$$

$$3Q_1 + 6Q_2 = 70$$

因此，单位时间的产出水平为

$$(\bar{Q}_1, \bar{Q}_2) = (8, 7\frac{2}{3})$$

356 分别将此结果代入(11.30′)和(11.31)，我们求得

$$\bar{P}_1 = 39\frac{1}{3} \quad \bar{P}_2 = 46\frac{2}{3} \quad \text{和} \quad \bar{\pi} = 488\frac{1}{3} \quad (\text{单位时间})$$

由于海塞矩阵为$\begin{vmatrix} -4 & -3 \\ -4 & -6 \end{vmatrix}$，我们有

$$|H_1| = -4 < 0, \qquad |H_2| = 15 > 0$$

所以$\bar{\pi}$ 值确定表示最大利润。这里，主子式的符号同样与其在何处计值无关。因此，海塞矩阵处处为负定，意味着目标函数为严格凹函数，且具有唯一的绝对极大值。

## 价格歧视

即便在单一产品厂商中，也会产生涉及两个或多个选择变量的最优化问题。比如，可能会有这种情况：一个垄断厂商在两个或多个隔离的(比如国内和国外)市场中销售单一产品，因此必须确定向每个市场分别供给的数量($Q_1$、$Q_2$ 等)以使利润最大化。一般而言，这几个不同的市场会有不同的需求条件，如果在不同市场中需求弹性不同，利润最大化就会涉及价格歧视问题。下面从数学上我们导出这个大家熟悉的结论。

**例 3** 为改变一下步调，这次我们选择三个选择变量，即假设

存在三个隔离的市场。我们仍使用一般函数，而不使用数字函数。因此，我们仅假定此厂商具有如下总收益函数和总成本函数：

$$R = R_1(Q_1) + R_2(Q_2) + R_3(Q_3)$$

$$C = C(Q) \quad \text{其中 } Q = Q_1 + Q_2 + Q_3$$

注意，这里符号 $R_i$ 表示第 $i$ 个市场的收益函数，它并没有 $f_i$ 这种导数意义。每个这样的收益函数自然意味着特殊的需求结构，它与另外两个市场流行的需求结构一般有所不同。另外，在成本方面，假设仅有一个成本函数，因为一个厂商为所有三个市场供应产品。考虑到 $Q = Q_1 + Q_2 + Q_3$，总成本 $C$ 基本上也是模型中三个选择变量 $Q_1$、$Q_2$、$Q_3$ 的函数。当然，我们可以将 $C(Q)$ 重写成 $C(Q_1 + Q_2 + Q_3)$。但应注意，尽管后一种形式包含三个自变量，但此函数仍可视为仅有一个变数，因为 $Q_i$ 的和实际上是一个整体。相反，若函数以 $C(Q_1, Q_2, Q_3)$ 这种形式出现，则出现几个变数就应视作有几个自变量。

357

现在利润函数为

$$\pi = R_1(Q_1) + R_2(Q_2) + R_3(Q_3) - C(Q)$$

其一阶偏导数 $\pi_i \equiv \partial\pi/\partial Q_i$（对于 $i = 1,2,3$）如下：[①]

$$\pi_1 = R'_1(Q_1) - C'(Q)\frac{\partial Q}{\partial Q_1}$$

$$= R'_1(Q_1) - C'(Q) \qquad \left[\text{因 } \frac{\partial Q}{\partial Q_1} = 1\right]$$

$$\pi_2 = R'_2(Q_2) - C'(Q)\frac{\partial Q}{\partial Q_2}$$

---

① 要求出 $\partial C/\partial Q_i$，可应用链式法则：$\frac{\partial C}{\partial Q_i} = \frac{dC}{dQ}\cdot\frac{\partial Q}{\partial Q_i}$。

$$= R'_2(Q_2) - C'(Q) \qquad [\text{因} \frac{\partial Q}{\partial Q_2} = 1] \quad (11.32)$$

$$\pi_3 = R'_3(Q_3) - C'(Q)\frac{\partial Q}{\partial Q_3}$$

$$= R'_3(Q_3) - C'(Q) \qquad [\text{因} \frac{\partial Q}{\partial Q_3} = 1]$$

令上述方程等于零，同时得到

$$C'(Q) = R'_1(Q_1) = R'_2(Q_2) = R'_3(Q_3)$$

即

$$\mathrm{MC} = \mathrm{MR}_1 = \mathrm{MR}_2 = \mathrm{MR}_3$$

因此，所选择的供给量水平 $Q_1$、$Q_2$、$Q_3$ 应使得每个市场的边际收益等于总产出 $Q$ 的边际成本。

为理解此条件关于价格歧视的含义，我们首先找出任意市场的 MR 是如何具体地与那个市场的价格联系起来的。因为每个市场的收益是 $R_i = P_iQ_i$，由此可知，边际收益必然为

$$\mathrm{MR}_i \equiv \frac{dR_i}{dQ_i} = P_i\frac{dQ_i}{dQ_i} + Q_i\frac{dP_i}{dQ_i}$$

$$= P_i(1 + \frac{dP_i}{dQ_i}\frac{Q_i}{P_i}) = P_i(1 + \frac{1}{\varepsilon_{di}}) \quad [\text{由 } (8.4)]$$

其中 $\varepsilon_{di}$ 为第 $i$ 个市场的点弹性，通常为负。因此，$\mathrm{MR}_i$ 与 $P_i$ 之间的关系可分别以如下方程表示

$$\mathrm{MR}_i = P_i\left(1 - \frac{1}{|\varepsilon_{di}|}\right) \qquad (11.33)$$

回忆一下，$|\varepsilon_{di}|$ 一般是 $P_i$ 的函数，因此当 $\bar{Q}_i$ 选定时，$P_i$ 便确定了，$|\varepsilon_{di}|$ 也将取定为一具体的值，它或者大于 1，或者等于 1，或者小于 1。但若 $|\varepsilon_{di}| < 1$（需求在该点缺乏弹性），则其倒数大于 1，

(11.33)中括号中的式子为负,因而意味着 $MR_i$ 的值为负。类似 358
地,若$|\varepsilon_{di}|=1$(单位单弹),则 $MR_i$ 将取零值。由于厂商的 MC 为正,一阶条件 $MC=MR_i$ 要求厂商在 $MR_i$ 为正的水平上经营,因此厂商所选择的销售水平 $Q_i$ 必须使该市场中相对应的点弹性大于 1。

根据(11.33),一阶条件 $MR_1=MR_2=MR_3$ 现在可以变换成如下形式:

$$P_1\left(1-\frac{1}{|\varepsilon_{d1}|}\right)=P_2\left(1-\frac{1}{|\varepsilon_{d2}|}\right)=P_3\left(1-\frac{1}{|\varepsilon_{d3}|}\right)$$

由此式可很容易推断出:在某一特定市场中(在选定之产出水平),$|\varepsilon_d|$越小,在该市场中所要索取的价格必须越高——即价格歧视——如果要使利润最大化。

为确保最大化,我们检验二阶条件。由(11.32),求得二阶偏导数如下

$$\pi_{11}=R''_1(Q_1)-C''(Q)\frac{\partial Q}{\partial Q_1}=R''_1(Q_1)-C''(Q)$$

$$\pi_{22}=R''_2(Q_2)-C''(Q)\frac{\partial Q}{\partial Q_2}=R''_2(Q_2)-C''(Q)$$

$$\pi_{33}=R''_3(Q_3)-C''(Q)\frac{\partial Q}{\partial Q_3}=R''_3(Q_3)-C''(Q)$$

和 $\pi_{12}=\pi_{21}=\pi_{13}=\pi_{31}=\pi_{23}=\pi_{32}=-C''(Q)$

$$\left[\text{因 }\frac{\partial Q}{\partial Q_i}=1\right]$$

所以在简化二阶导数符号后,我们有

$$|H|=\begin{vmatrix} R''_1-C'' & -C'' & -C'' \\ -C'' & R''_2-C'' & -C'' \\ -C'' & -C'' & R''_3-C'' \end{vmatrix}$$

因此，如果下述要求成立，则二阶充分条件便完全满足：

1. $|H_1| = R_1'' - C'' < 0$；即 $MR_1$ 的斜率小于整个产出 MC 的斜率[参见图 9.6c 中点 $L$ 的情形]。（因为三个市场中任意一个市场都可以视为“第一个”市场，所以这实际上也意味着 $R_2'' - C'' < 0$ 和 $R_3'' - C'' < 0$。）

2. $|H_2| = (R_1'' - C'')(R_2'' - C'') - (C'')^2 > 0$；或，$R_1''R_2'' - (R_1'' + R_2'')C'' > 0$

3. $|H_3| = R_1''R_2''R_3'' - (R_1''R_2'' + R_1''R_3'' + R_2''R_3'')C'' < 0$

上述条件的后两部分的经济意义并不像第一部分那样容易解释。注意，如果我们假设一般的 $R_i(Q_i)$ 函数均为凹函数，一般的 $C(Q)$ 函数为凸函数，那么，$-C(Q)$ 是凹函数，则利润函数——凹函数之和——也可以视为凹函数，从而避免了检验二阶条件的必要性。

359 **例 4** 为使上述例子更具体，我们现在给出一个数字形式的例子。假定我们的垄断厂商具有如下具体的平均收益函数

$$P_1 = 63 - 4Q_1 \quad \text{从而} \quad R_1 = P_1Q_1 = 63Q_1 - 4Q_1^2$$

$$P_2 = 105 - 5Q_2 \qquad R_2 = P_2Q_2 = 105Q_2 - 5Q_2^2$$

$$P_3 = 75 - 6Q_3 \qquad R_3 = P_3Q_3 = 75Q_3 - 6Q_3^2$$

且总成本函数为

$$C = 20 + 15Q$$

则边际函数为

$$R'_1 = 63 - 8Q_1 \quad R'_2 = 105 - 10Q_2 \quad R'_3 = 75 - 12Q_3 \quad C' = 15$$

当令每个边际收益 $R'_i$ 等于总产出的边际成本 $C'$ 时，可求得均衡数量为

$$\bar{Q}_1 = 6 \quad \bar{Q}_2 = 9 \quad 和 \quad \bar{Q}_3 = 5$$

则 $$\bar{Q} = \sum_{i=1}^{3} \bar{Q}_i = 20$$

将上述结果代入收益和成本方程,我们得到$\bar{\pi} = 679$,这是在三个市场经营中得到的总利润。

因为这是一个具体模型,我们必须检验二阶条件(或目标函数的凹性)。因为二阶导数为:

$$R''_1 = -8 \quad R''_2 = -10 \quad R''_3 = -12 \quad C'' = 0$$

所以,例 3 中给定的二阶充分条件的三部分内容完全得到满足。

由平均收益函数很容易看出:厂商在三个市场中应索取不同的价格 $\bar{P}_1 = 39, \bar{P}_2 = 60, \bar{P}_3 = 45$。正如你已经验证的那样,在第二个市场中需求的点弹性最低,所索要的价格也最高。

## 厂商的投入决策

除了产出水平 $Q_i$ 外,厂商的选择变量还可以是投入水平。

**例 5** 我们假设存在如下环境:(1) 一个假定的厂商运用投入 $a$ 和 $b$ 生产单一产品 $Q$。(2) 投入的 $P_a$ 和 $P_b$ 不能为该厂商所控制,该厂商也不能控制其产出的价格 $\bar{P}$,所以我们将其分别表示 360
成 $P_{a0}, P_{b0}, P_0$。(3) 生产过程要 $t_0$ 年($t_0$ 为正常数)完成,所以在将销售收益与现在发生的投入成本进行比较之前,需将其完全贴现至现值。在连续基础上,假定贴现率为 $r_0$。

根据假设 1,我们可以写出一般生产函数 $Q = Q(a, b)$,其边际物质产品为 $Q_a, Q_b$。由假设 2,可将总成本函数表示成

$$C = aP_{a0} + bP_{b0}$$

将总收益函数表示成

$$R = P_0 Q(a, b)$$

为写出利润函数，我们必须将收益与一个常数 $e^{-r_0 t_0}$ 相乘以将收益折现，为避免上标与下标混淆，我们将 $e^{-r_0 t_0}$ 写成 $e^{-rt}$。因此，利润函数为

$$\pi = P_0 Q(a, b) e^{-rt} - aP_{a0} - bP_{b0}$$

其中仅有 $a$ 和 $b$ 为选择变量。

为使利润最大化，必须先求出一阶偏导数：

$$\begin{aligned} \pi_a \left(\equiv \frac{\partial \pi}{\partial a}\right) &= P_0 Q_a e^{-rt} - P_{a0} \\ \pi_b \left(\equiv \frac{\partial \pi}{\partial b}\right) &= P_0 Q_b e^{-rt} - P_{b0} \end{aligned} \tag{11.34}$$

令二者均为零，这意味着

$$P_0 Q_a e^{-rt} = P_{a0} \quad \text{和} \quad P_0 Q_b e^{-rt} = P_{b0} \tag{11.35}$$

因为 $P_0 Q_a$（产品价格乘投入 $a$ 的边际产品）表示**投入 $a$ 的边际产品价值**（$VMP_a$），所以第一个方程仅表示 $VMP_a$ 的现值应等于投入 $a$ 的给定价格。第二个方程表示同样的先决条件应用于投入 $b$。

注意，要满足(11.35)，边际物质产品 $Q_a$ 和 $Q_b$ 必须均为正，因为 $P_0, P_{a0}, P_{b0}$，以及 $e^{-rt}$ 均为正值。按照**等产量曲线**，它具有非常重要的解释；等产量曲线定义为能获得同样产出水平的投入组合的轨迹。当在 $ab$ 平面上绘出时，等产量曲线一般与图 11.11 中出现的曲线类似。由于每一等产量曲线代表一个固定产出水平，在任意等产量曲线上，我们必然有

$$dQ = Q_a da + Q_b db = 0$$

这意味着等产量曲线的斜率可以表达成

$$\frac{db}{da}=-\frac{Q_a}{Q_b}\quad\left(=-\frac{\mathrm{MPP}_a}{\mathrm{MPP}_b}\right)\tag{11.36}$$

因此,要使 $Q_a$ 和 $Q_b$ 均为正,必须把厂商的投入选择限定在等产

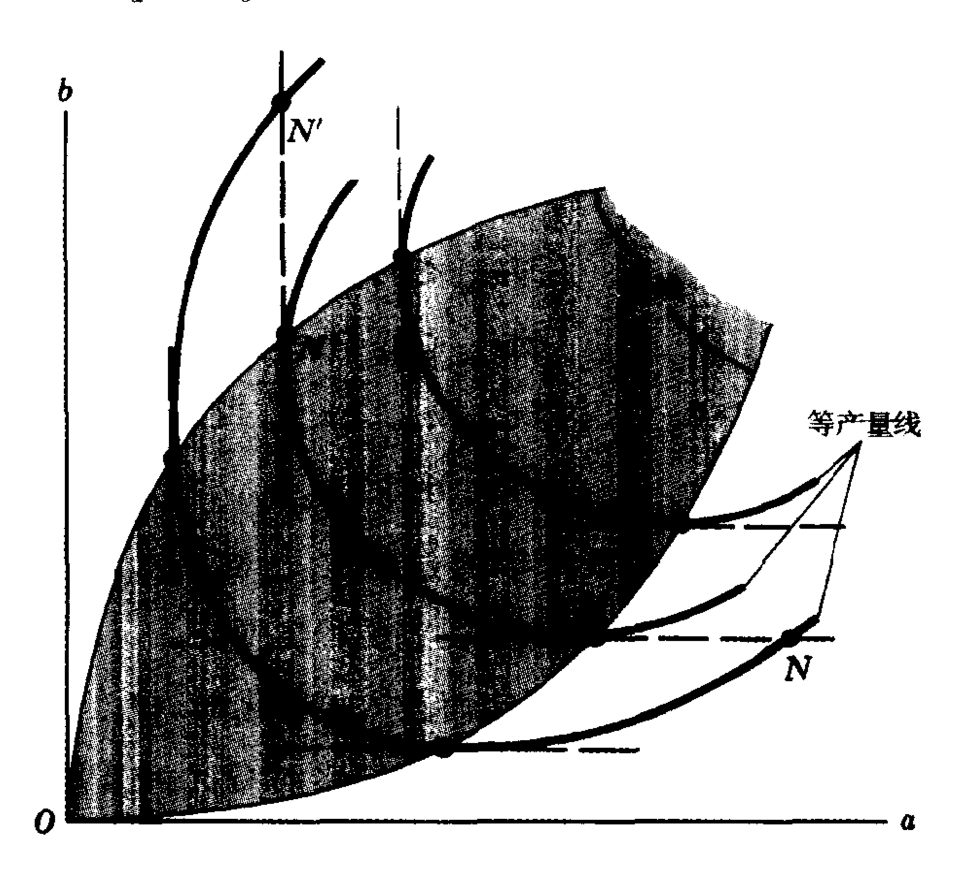



图 11.11

量曲线斜率为负的弧段内。在图 11.11 中,经营的相关领域是由两条所谓的“脊线”所确定的阴影区域。在阴影区域外,等产量曲线的斜率为正,一种投入的边际产品必定为负。例如,由投入组合 $M$ 移至 $N$,表明投入 $b$ 保持不变时增加投入 $a$,使我们达到较低的等产量线(一个较小的产出),因此,$Q_a$ 必定为负。类似地,$M'$ 移至 $N'$ 表明 $Q_b$ 为负。注意,当我们将注意力集中于阴影区域时,每一等产量线可以视为形式为 $b=\phi(a)$ 的函数,因为对于每一个可接受的 $a$ 值,等产量曲线确定了一个唯一的 $b$ 值。 361

二阶条件由 $\pi$ 的二阶偏导数决定,而 $\pi$ 的二阶偏导数可以从

(11.34)求得。记住,作为导数,$Q_a$ 和 $Q_b$ 本身是变量 $a$ 和 $b$ 的函数,我们可以求得$\pi_{aa}$,$\pi_{ab}=\pi_{ba}$和$\pi_{bb}$,整理,列成海塞矩阵:

$$|H|=\begin{vmatrix}\pi_{aa} & \pi_{ab}\\ \pi_{ab} & \pi_{bb}\end{vmatrix}=\begin{vmatrix}P_0Q_{aa}e^{-rt} & P_0Q_{ab}e^{-rt}\\ P_0Q_{ab}e^{-rt} & P_0Q_{bb}e^{-rt}\end{vmatrix} \tag{11.37}$$

要使$\pi$ 的稳定值为极大值,其充分条件是

$|H_1|<0$ [即$\pi_{aa}<0$,当且仅当 $Q_{aa}<0$ 时,可得到$\pi_{aa}<0$。]

$|H_2|=|H|>0$ [即$\pi_{aa}\cdot\pi_{bb}>\pi_{ab}^2$,当且仅当 $Q_{aa}Q_{bb}>Q_{ab}^2$时,可得到$\pi_{aa}\pi_{bb}>\pi_{ab}^2$]

由此我们知道,二阶条件既可以通过导数 $\pi_{ij}$ 也可以通过导数 $Q_{ij}$ 来检验;至于在检验中运用哪个,视方便程度而定。

符号 $Q_{aa}$ 表示当投入$b$ 固定而投入$a$ 变化时,$Q_a$($\equiv\text{MPP}_a$)的变化率;类似地,$Q_{bb}$表示仅当投入 $b$ 变化时 $Q_b$($\equiv\text{MPP}_b$)的变化率。所以二阶充分条件部分地规定了两种投入的 MPP 在选定投入水平 $\bar{a}$ 和 $\bar{b}$ 是递减的。但要注意到,$\text{MPP}_a$ 和 $\text{MPP}_b$ 的递减并不能确保二阶条件的满足,因为二阶条件还包括 $Q_{ab}=Q_{bb}$的大小,它度量当一种投入的数量变化时,另一投入的 MPP 的变化率。

通过进一步考察可以揭示出,正如一阶条件在选定的投入组合(如图 11.11 阴影区域所示)中确定了等产量曲线斜率为负一样,二阶充分条件在选定的投入产出组合中确定了同一等产量曲线为严格凸。等产量曲线的曲率与二阶导数 $d^2b/da^2$ 的符号有关。要求得此二阶导数,必须将(11.36)对 $a$ 求全微分,记住 $Q_a$ 和 $Q_b$ 均为$a$ 与 $b$ 的函数,且在等产量曲线上,$b$ 本身是$a$ 的函数,

即

$$Q_a = Q_a(a, b) \quad Q_b = Q_b(a, b) \quad 和 \quad b = \phi(a)$$

因此可求全微分如下：

$$\frac{d^2 b}{da^2} = \frac{d}{da}\left(-\frac{Q_a}{Q_b}\right) = -\frac{1}{Q_b^2}\left[Q_b \frac{dQ_a}{da} - Q_a \frac{dQ_b}{da}\right] \tag{11.38}$$

因在等产量曲线上，$b$ 为 $a$ 的函数，由全导数公式(8.9)可知：

$$\frac{dQ_a}{da} = \frac{\partial Q_a}{\partial b}\frac{db}{da} + \frac{\partial Q_a}{\partial a} = Q_{ba}\frac{db}{da} + Q_{aa}$$

$$\frac{dQ_b}{da} = \frac{\partial Q_b}{\partial b}\frac{db}{da} + \frac{\partial Q_b}{\partial a} = Q_{bb}\frac{db}{da} + Q_{ab} \tag{11.39}$$

将(11.36)代入(11.39)后，再将后者代入(11.38)，则我们可以将二阶导数重写成

$$\begin{aligned}\frac{d^2 b}{da^2} &= -\frac{1}{Q_b^2}\left[Q_{aa}Q_b - Q_{ba}Q_a - Q_{ab}Q_a + Q_{bb}Q_a^2\left(\frac{1}{Q_b}\right)\right] \\ &= -\frac{1}{Q_b^3}\left[Q_{aa}(Q_b)^2 - 2Q_{ab}(Q_a)(Q_b) + Q_{bb}(Q_a)^2\right]\end{aligned} \tag{11.40}$$

应注意到，第二行括号中的表达式是两个变量 $Q_a$ 和 $Q_b$ 的二次型。若二阶充分条件得到满足，那么

$$Q_{aa} < 0 \quad 和 \quad \begin{vmatrix} Q_{aa} & -Q_{ab} \\ -Q_{ab} & Q_{bb} \end{vmatrix} > 0$$

则根据(11.11′)，上面所说的二次型必然为负定。这又将使 $d^2b/da^2$ 为正，因为 $Q_b$ 已为一阶条件限定为正。因此，二阶充分条件的满足意味着相关的(斜率为负的)等产量曲线在选定的投入组合中为严格凸，这与我们前面的结论一致。

363 严格凸性的概念应用于等产量曲线 $b=\phi(a)$时，应与将此概念应用于生产函数 $Q(a,b)$作严格的区分。$b=\phi(a)$是在二维 $ab$ 平面中绘出的，而 $Q(a,b)$则是在三维 $abQ$ 空间中绘出的。特别要注意，如果我们把严格凹性或严格凸性的概念应用于这里的生产函数，以便产生所需要的等产量曲线形状，适当的规定是 $Q(a,b)$在三维空间中为**严格凹**(呈山丘形)，这与规定相关的等产量曲线在二维空间中为**严格凸**(呈 U 形或 U 形的一部分)形成了鲜明的对照。

**例 6** (接例 5)其次，假设利息按季度计算复利，已知每季度利率为 $i_0$。还假设生产过程恰好需一季度完成。利润函数则变成

$$\pi=P_0Q(a,b)(1+i_0)^{-1}-aP_{a0}-bP_{b0}$$

现在求得一阶条件为

$$P_0Q_a(1+i_0)^{-1}-P_{a0}=0$$

$$P_0Q_b(1+i_0)^{-1}-P_{b0}=0$$

除了贴现方式不同外，其分析解释与例 5 完全相同。

读者可能已经看到，上一例中推导的充分条件也需要在这里应用。

## 练习 11.6

---

1 若例 1 中的竞争厂商拥有成本函数 $C=2Q_1^2+2Q_2^2$，则

($a$) 两产品的生产仍存在技术相关性吗？

($b$) $Q_1$ 和 $Q_2$ 的新的最优水平是多少？

（$c$）$\pi_{12}$的值是多少？如何解释其经济含义？

2　一个两产品厂商的需求函数和成本函数如下：

$$Q_1 = 40 - 2P_1 - P_2 \quad Q_2 = 35 - P_1 - P_2 \quad C = Q_1^2 + 2Q_2^2 + 10$$

（$a$）求满足利润最大化一阶条件的产出水平。(结果用分数表示。)

（$b$）检验二阶充分条件。你能断定此题只有一个绝对极大值吗？

（$c$）最大利润为多少？

3　在例 4 均衡价格和均衡利润的基础上，计算需求的点弹性$|\varepsilon_{di}|$($i=1,2,3$)。哪个市场的需求弹性最高？哪个市场的需求弹性最低？

4　若例 4 的成本函数变为 $C = 20 + 15Q + Q^2$：　364

（$a$）求此新的边际成本函数。

（$b$）求新的均衡数量。(用分数表示结果。)

（$c$）求新的均衡价格。

（$d$）验证二阶充分条件是满足的。

5　在例 6 中，若下列条件成立，你如何重新写出利润函数？

（$a$）利息以年利率 $i_0$ 按半年计算复利，且生产过程为 1 年。

（$b$）利息以年利率 $i_0$ 按季度计算复利，生产过程为 9 个月。

6　给定 $Q = Q(a,b)$，你如何用代数形式表示等产量曲线。

---

## 11.7　最优化的比较静态方面

最优化作为一种特殊类型的比较静态分析，自然也可以用于研究比较静态方面的问题。其思想仍然是求出任意参数的变化将如何影响模型的均衡状态；在这里模型的均衡状态是指选择变量的最优值(以及目标函数的最优值)。除了在第三篇中讨论的内容之外，这里并不涉及新的技术，我们在上节介绍的例子的基础上，直接展开陈述。

## 简化型解

11.6 节中的例 1 包含两个参数(或外生变量)$P_{10}$和 $P_{20}$,因此,毫不奇怪,这个两产品厂商的最优产出水平可严格地按照这两个参数表示:

$$\bar{Q}_1=\frac{4P_{10}-P_{20}}{15} \quad \text{和} \quad \bar{Q}_2=\frac{4P_{20}-P_{10}}{15}$$

这两个是简化型解,简单的偏微分便足以告诉我们模型的比较静态性质,即

$$\frac{\partial\bar{Q}_1}{\partial P_{10}}=\frac{4}{15} \qquad \frac{\partial\bar{Q}_1}{\partial P_{20}}=-\frac{1}{15} \qquad \frac{\partial\bar{Q}_2}{\partial P_{10}}=-\frac{1}{15} \qquad \frac{\partial\bar{Q}_2}{\partial P_{20}}=\frac{4}{15}$$

要获得最大利润,当某产品价格上升或另一种产品价格下降时,厂商应扩大该产品的产量。

当然,这些结论仅在所研究模型的特定假设下才成立。我们特别应当指出,$P_{10}$变化对 $\bar{Q}_2$ 的影响、$P_{20}$变化对 $\bar{Q}_1$ 的影响是假设两产品在生产方面存在技术相关性之结果。在不存在这种技术相关性时,我们有

365

$$\frac{\partial\bar{Q}_1}{\partial P_{20}}=\frac{\partial\bar{Q}_2}{\partial P_{10}}=0$$

现在考察例 2,我们注意到最优产出水平是以数字表述的,$\bar{Q}_1=8,\bar{Q}_2=7\frac{2}{3}$——没有参数出现。事实上,模型方程所有的常数均是数字而非参数的,所以当我们进入求解阶段时,经过算术运算过程,这些常数均已失去了其各自的特性。这表明由于使用数字常数缺乏一般性,所以均衡解的结果也缺乏比较静态含义。

另一方面,不使用数字常数就不能保证一个问题能自然符合

比较静态分析的要求。比如,价格歧视问题(例3)基本上是为研究均衡(利润最大化)条件而设立的,没有引进任何参数。相应地尽管按照一般函数形式表述,若比较静态研究需要,重新按照公式表述也是必要的。

## 一般函数模型

例5中的投入决策问题表明一般函数公式确实包含了几个参数——事实上不少于五个($P_0,P_{a0},P_{b0},r$ 和 $t$),这里我们同以前一样,省略了外生变量 $r_0$ 和 $t_0$ 的下标。我们如何推导此模型的比较静态性质呢?

答案仍在于应用隐函数定理。但是,这次与市场或国民收入决定的非目标均衡不同,在现在的目标均衡中,我们利用最优化的一阶条件,而在非目标均衡中,我们利用模型的均衡条件。对于例5而言,这些条件表述在(11.35)中。将(11.35)中各项移至等号左边,并使 $Q_a$ 和 $Q_b$ 更明显地表示成内生(选择)变量 $a$ 和 $b$ 的函数,我们可以将在(8.20)中的一阶条件重写成如下形式:

$$F^1(a,b;P_0,P_{a0},P_{b0},r,t)=P_0Q_a(a,b)e^{-rt}-P_{a0}=0$$

$$F^2(a,b;P_0,P_{a0},P_{b0},r,t)=P_0Q_b(a,b)e^{-rt}-P_{b0}=0 \tag{11.41}$$

假设函数 $F^1$ 和 $F^2$ 拥有连续导数。如果此方程组对内生变量 $a$、$b$ 的雅可比行列式在初始均衡时不为零,则可以应用隐函数定理。这里所说的雅可比行列式只不过是例5中 $\pi$ 函数的海塞行列式: 366

$$|J|=\begin{vmatrix}\dfrac{\partial F^1}{\partial a} & \dfrac{\partial F^1}{\partial b}\\[2ex] \dfrac{\partial F^2}{\partial a} & \dfrac{\partial F^2}{\partial b}\end{vmatrix}=\begin{vmatrix}P_0Q_{aa}e^{-rt} & P_0Q_{ab}e^{-rt}\\ P_0Q_{ab}e^{-rt} & P_0Q_{bb}e^{-rt}\end{vmatrix}=|H|$$

$$[\text{由}(11.37)] \qquad (11.42)$$

因此,若我们假设利润最大化的二阶充分条件得到满足,则在初始均衡或极值处,$|H|$必定为正,$|J|$也可正。在这种情况下,根据隐函数定理,我们可以写出以下两个隐函数

$$\bar{a} = \bar{a}\ (P_0, P_{a0}, P_{b0}, r, t)$$
$$\bar{b} = \bar{b}\ (P_0, P_{a0}, P_{b0}, r, t) \qquad (11.43)$$

以及以下两个恒等式

$$P_0 Q_a(\bar{a}, \bar{b}) e^{-rt} - P_{a0} \equiv 0$$
$$P_0 Q_b(\bar{a}, \bar{b}) e^{-rt} - P_{b0} \equiv 0 \qquad (11.44)$$

为研究此模型的比较静态性质,首先取(11.44)中每个恒等式的全微分。暂时我们允许外生变量变化,所以全微分的结果将包含 $d\bar{a}$,$d\bar{b}$,以及 $dP_0, dP_{a0}, dP_{b0}, dr$ 和 $dt$。如果我们将等号左边仅置放那些包含 $d\bar{a}$ 和 $d\bar{b}$ 的项,则结果将为

$$P_0 Q_{aa} e^{-rt} d\bar{a} + P_0 Q_{ab} e^{-rt} d\bar{b}$$
$$= -Q_a e^{-rt}\, dP_0 + dP_{a0} + P_0 Q_a t e^{-rt}\, dr + P_0 Q_a r e^{-rt} dt$$
$$P_0 Q_{ab} e^{-rt} d\bar{a} + P_0 Q_{bb} e^{-rt} d\bar{b}$$
$$= -Q_b e^{-rt}\, dP_0 + dP_{b0} + P_0 Q_b t e^{-rt}\, dr + P_0 Q_b r e^{-rt} dt \qquad (11.45)$$

应注意,$Q$ 的一阶和二阶导数均在均衡处,亦即在 $\bar{a}$ 和 $\bar{b}$ 处计值。读者还要注意 $d\bar{a}$ 和 $d\bar{b}$ 左边的系数正好是(11.42)中雅可比行列式的元素。

为导出具体的比较静态导数(共有 10 个,为什么?),我们现在

每次仅允许一个外生变量发生变化。假设我们仅允许 $P_0$ 变化，则 $dP_0 \neq 0$，但 $dP_{a0} = dP_{b0} = dr = dt = 0$，所以在(11.45)中每个方程的右边剩下一项。以 $dP_0$ 通除，并将比率 $d\bar{a}/dP_0$ 解释为比较静态导数$(\partial\bar{a}/\partial P_0)$，并将比率 $d\bar{b}/dP_0$ 也作类似解释，则我们可以解出矩阵方程

$$\begin{bmatrix} P_0 Q_{aa} e^{-rt} & P_0 Q_{ab} e^{-rt} \\ P_0 Q_{ab} e^{-rt} & P_0 Q_{bb} e^{-rt} \end{bmatrix} \begin{bmatrix} (\partial\bar{a}/\partial P_0) \\ (\partial\bar{b}/\partial P_0) \end{bmatrix} = \begin{bmatrix} -Q_a e^{-rt} \\ -Q_b e^{-rt} \end{bmatrix}$$

由克莱姆法则，求得解为

367

$$\left(\frac{\partial\bar{a}}{\partial P_0}\right) = \frac{(Q_b Q_{ab} - Q_a Q_{bb}) P_0 e^{-2rt}}{|J|}$$
$$\left(\frac{\partial\bar{b}}{\partial P_0}\right) = \frac{(Q_a Q_{ab} - Q_b Q_{aa}) P_0 e^{-2rt}}{|J|} \tag{11.46}$$

如果你愿意，还可以用其它方法得到这些结果：你可以将(11.44)中的两个恒等式对 $P_0$ 求全微分（使其它四个外生变量保持不变），但要记住 $P_0$ 可通过(11.43)影响$\bar{a}$和$\bar{b}$。

现在我们来分析(11.46)中比较静态导数的符号。在二阶充分条件满足的假设下，分母中的雅可比行列式必定为正。正如一阶条件意味着 $Q_a$ 和 $Q_b$ 为正一样，二阶条件也意味着 $Q_{aa}$和 $Q_{bb}$ 为负。而且，表达式 $P_0 e^{-2rt}$必然为正，因此，若 $Q_{ab} > 0$（若增加一种投入将提高另一投入的 MPP），我们可以得出结论：$(\partial\bar{a}/\partial P_0)$ 和$(\partial\bar{b}/\partial P_0)$均为正，这意味着产品价格的提高将导致均衡时两种投入使用量的增加。另一方面，若 $Q_{ab} < 0$，(11.46)中每个导数的符号将取决于右边括号中表达式的正项与负项的相对大小。

其次，仅令外生变量 $r$ 变化。则(11.45)中右边各项除了含

$dr$ 的项以外，均为零。以 $dr \neq 0$ 通除，我们现在可得到如下矩阵方程：

$$\begin{bmatrix} P_0 Q_{aa} e^{-rt} & P_0 Q_{ab} e^{-rt} \\ P_0 Q_{ab} e^{-rt} & P_0 Q_{bb} e^{-rt} \end{bmatrix} \begin{bmatrix} (\partial \bar{a} / \partial r) \\ (\partial \bar{b} / \partial r) \end{bmatrix} = \begin{bmatrix} P_0 Q_a t e^{-rt} \\ P_0 Q_b t e^{-rt} \end{bmatrix}$$

其解为

$$\left( \frac{\partial \bar{a}}{\partial \mathrm{r}} \right) = \frac{t(Q_a Q_{bb} - Q_b Q_{ab})(P_0 e^{-rt})^2}{|J|}$$

$$\left( \frac{\partial \bar{b}}{\partial \mathrm{r}} \right) = \frac{t(Q_b Q_{aa} - Q_a Q_{ab})(P_0 e^{-rt})^2}{|J|} \qquad (11.47)$$

若 $Q_{ab}$ 为正，这两个比较静态导数为负，但若 $Q_{ab}$ 为负，则其符号不确定。

同类似的步骤，我们可以求出余下参数变化的影响。实际上，考虑到(11.44)中 $r$ 和 $t$ 的对称性，立即就可以明了 $(\partial \bar{a} / \partial t)$ 和 $(\partial \bar{b} / \partial t)$ 必然会有与(11.47)相类似的形式。

$P_{a0}$ 和 $P_{b0}$ 变化的影响留给读者自己分析。正如你将发现的那样，二阶充分条件的符号限制在计算比较静态导数时仍然是有用的，因为它可以告诉我们 $Q_{aa}$、$Q_{bb}$ 的符号，以及在初始均衡(最优值)处雅可比行列式 $|J|$ 的符号。因此，除了区分极大值和极小值外，二阶条件在均衡状态移动的研究中，也发挥非常重要的作用。

368

## 练习 11.7

---

对于如下三个问题，假设 $Q_{ab} > 0$。

1 在(11.41)至(11.44)所描述模型的基础上，求比较静态导数 $(\partial \bar{a} /$

$\partial P_{a0}$)和($\partial \bar{b}/\partial P_{a0}$)。解释此结果的经济意义。并分析 $P_{b0}$变化对 $\bar{a}$ 和 $\bar{b}$ 的影响。

2 对于11.6节中例6的问题：

($a$) 其中有多少个参数？请列出每个参数。

($b$) 按照(11.41)至(11.46)所描述的步骤，并假设满足二阶充分条件，求比较静态导数($\partial \bar{a}/\partial P_0$)和($\partial \bar{b}/\partial P_0$)。判断其符号并解释其经济意义。

($c$) 求($\partial \bar{a}/\partial i_0$)和($\partial \bar{b}/\partial i_0$)，判断其符号，解释其经济含义。

3 证明结果(11.46)还可以通过将(11.44)中的恒等式对 $P_0$ 求全微分，同时保持其它外生变量不变的方法得到。记住 $P_0$ 由于(11.43)能够对 $\bar{a}$ 和 $\bar{b}$ 产生影响。

4 一个雅可比行列式，正如(7.27)所定义那样，由一阶导数构成。而某海塞行列式，如11.3节和11.4节所定义那样，其元素则是二阶偏导数，那么，如何才能像(11.42)那样，证明 $|J| = |H|$？

369

# 第12章 具有约束方程的最优化

上一章提供了求具有两个或多个选择变量目标函数相对极值的一般方法。这些讨论的一个重要特征是所有的选择变量彼此独立无关,在这个意义上,关于一个选择变量的决策不会对其它变量产生任何影响。例如,一个两产品厂商可以选择他所期望的任意 $Q_1$ 值和 $Q_2$ 值,这两种选择不会彼此限制。

但是,如果上面所提到的厂商需要遵守形式为 $Q_1+Q_2=950$ 的约束(比如生产配额),那么,选择变量间的独立性便不存在了。在这种情况下,厂商利润最大化的产出水平 $\bar{Q}_1$ 和 $\bar{Q}_2$ 不仅同时确定,而且还彼此依赖,因为要维持在950的配额内,$\bar{Q}_1$ 提高,$\bar{Q}_2$ 便要相应降低。这种新的需要满足生产配额的最优化称作约束最优化,它一般有别于上一章所讨论的自由最优化。

一个约束,比如上面提到的生产配额,确立了两个变量在充当选择变量时的关系,但这种关系与其它类型的将两个变量联系在一起的关系是不同的。比如,在11.6节例2中,厂商的两种产品 370 在生产(反映在成本函数中)和消费(替代关系)方面存在着相互联系,但这并不是约束最优化问题,因为两个产出变量作为选择变量依然是独立无关的。

在本章,我们将仅考察等式约束,比如 $Q_1+Q_2=950$ 等。我们主要关注的是相对约束极值,尽管在12.4节也讨论绝对极值。

## 12.1 约束的影响

施加约束的基本目的是对在所讨论的最优化问题中存在的某些限制因素给出合理的认识。

我们已经看到对产出选择的限制会导致生产配额。为进一步对此进行解释,我们考察一个具有简单效用函数的消费者

$$U = x_1x_2 + 2x_1 \tag{12.1}$$

因为边际效用——偏导数 $U_1 \equiv \partial U/\partial x_1$ 和 $U_2 \equiv \partial U/\partial x_2$——对所有的正的 $x_1$ 和 $x_2$ 均为正,要使 $U$ 最大化而不施加任何约束,该消费者需购买无穷多数量的两种商品,此解显然没有任何实际意义。为使得此最优化问题有意义,我们必须考虑消费者的购买力问题,即把预算约束纳入问题之中。如果消费者在两种商品上准备支付的总额为 60 美元,两种商品价格为 $P_{10} = 4$ 和 $P_{20} = 2$,则预算约束可以用线性方程表示:

$$4x_1 + 2x_2 = 60 \tag{12.2}$$

这个约束,同前面提到的生产配额一样,使得 $\bar{x}_1$ 和 $\bar{x}_2$ 的选择相互依赖。

现在的问题是在(12.2)约束下,使(12.1)最大化。从数学上看,约束(也称限制、边条件或附加条件)的作用在于缩小定义域,从而缩小目标函数的值域。(12.1)的定义域正常情况下为集合 $\{(x_1, x_2) \mid x_1 \geqslant 0, x_2 \geqslant 0\}$。从图形上看,此定义域是由图 12.1$a$ 中 $x_1x_2$ 平面中的非负象限表示的。但加上预算约束(12.2)以后,我们只能接受那些满足后一方程的变量值,所以定义域立即缩减为位于预算线上点的集合。这自然也会影响到目标函数的值域,

只有位于预算约束线正上方的效用曲面的子集，才是相关的值域。刚才提到的子集(曲面的截痕)看起来像图 12.1$b$ 中的曲线，其中 $U$ 标在纵轴上，而图 $a$ 中的预算线则绘在横轴上。那么，我们感兴趣的是确定图 $b$ 中曲线的极大值。

371

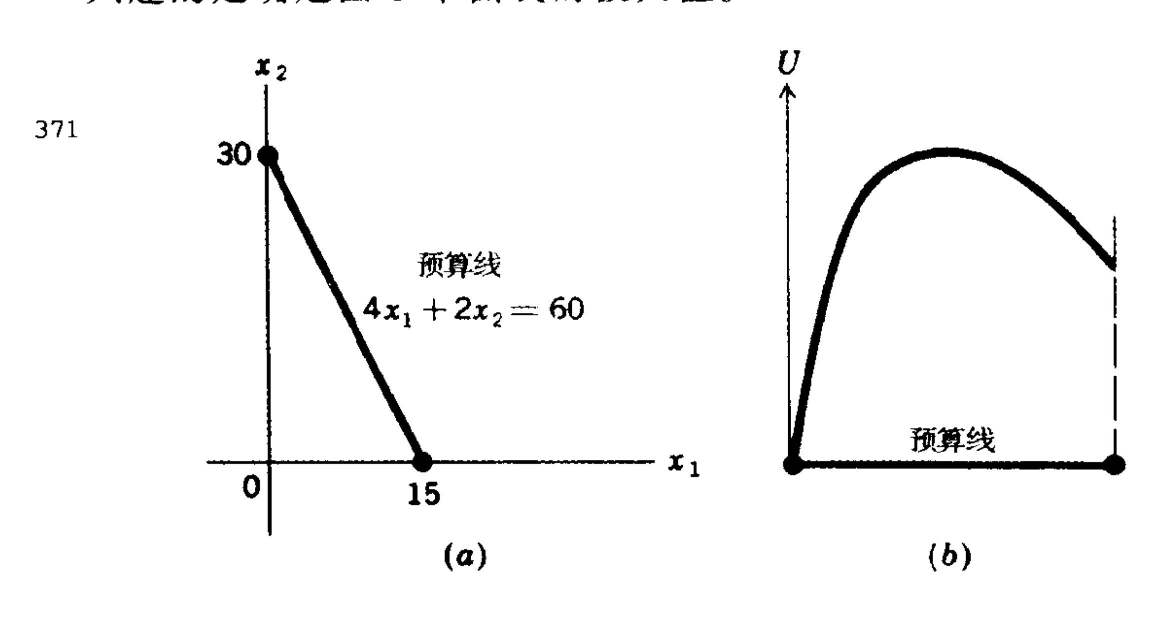



图 12.1

一般而言，对于函数 $z=f(x, y)$，约束极值与自由极值的差别可以通过图 12.2 中的三维图形来描述。在此特定图形中，自由极值是整个山丘的峰值，而约束极值是位于约束约正上方的倒 $U$ 形曲线的峰值。一般情况下，可以预期约束极值要小于自由极值，尽管偶然情况下它们也可能恰巧相等。但约束极值永远不能大于自由极值。

值得注意的是，如果我们增加另一个约束，此约束与第一个约束在 $xy$ 平面上相交于一个点，则这两个约束的结合使定义域限定在这个点上。因而确定极值变得极其简单。在一个有意义的问题中，约束的数量和性质应当是限制选择的可能性，而不是排除选择的可能性。一般而言，约束的数量应少于选择变量的数量。

372

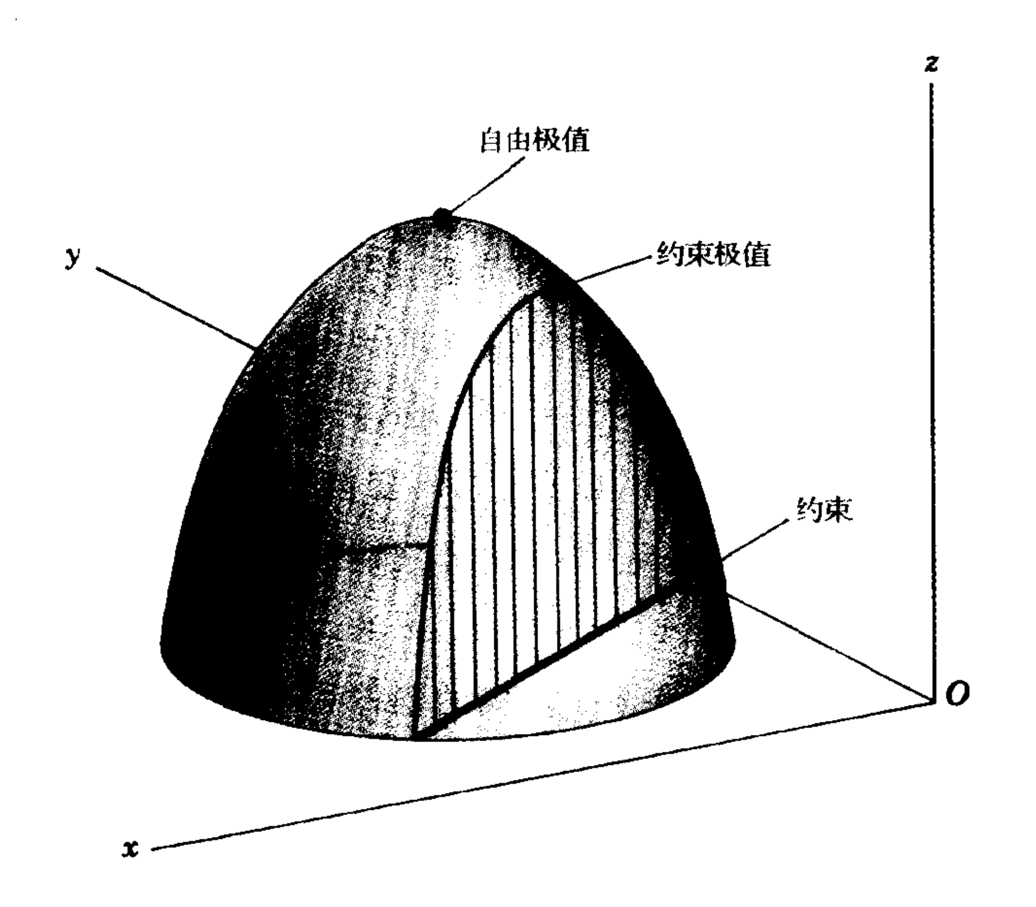



图 12.2

## 12.2 求稳定值

即使不运用新的求解方法，像在(12.1)和(12.2)中所定义的简单约束极值也可以轻松求出。因为约束(12.2)意味着

$$x_2 = \frac{60 - 4x_1}{2} = 30 - 2x_1 \tag{12.2'}$$

我们可以通过把(12.2′)代入(12.1)而将约束与目标函数结合起来。结果得到一个具有一元变量的目标函数：

$$U = x_1(30 - 2x_1) + 2x_1 = 32x_1 - 2x_1^2$$

它可以用早已知道的方法求解。令 $dU/dx_1 = 32 - 4x_1$ 等于零，我们可以得到解 $\bar{x}_1 = 8$，利用(12.2′)马上可以得到 $\bar{x}_2 = 30 - 2(8) = 14$。由(12.1)，我们可以得到稳定值 $\bar{U} = 128$，又因二阶导数 $d^2U/dx_1^2 = -4 < 0$，所以稳定值亦即(约束)极大值。[1]

但是，当约束函数本身是一个复杂的函数，或者存在几个需要考察的函数时，以代入法和消元法求解变得非常麻烦。更重要的是，当约束以这种形式出现：即我们不能解约束方程，从而将一个变量($x_2$)表示成另一个变量($x_1$)的显函数，那么，即使知道 $x_2$ 是 $x_1$ 的隐函数，亦即即使隐函数定理的条件得到满足，实际上也难以应用消元法。在这种情况下，我们可以借助于**拉格朗日乘数法**。正如我们将要看到的那样，拉格朗日乘数法具有别具一格的分析优势。

## 拉格朗日乘数法

拉格朗日乘数法的实质是将约束极值问题转化成这样一种形式，从而使得自由极值问题的一阶条件仍可以应用。

给定 $U = x_1x_2 + 2x_1$，约束条件 $4x_1 + 2x_2 = 60$[由(12.1)和(12.2)]，求 $U$ 的极大值。我们先写出所谓的**拉格朗日函数**，它是容纳了约束条件的目标函数的一种变形：

373

$$Z = x_1x_2 + 2x_1 + \lambda(60 - 4x_1 - 2x_2) \tag{12.3}$$

---

① 读者回忆一下，练习 9.4-2 中的花坛问题，使用过同样的代入法，即用约束条件(供使用的金属网数量)从两个变量(花坛的长度)消去一个变量，求出最大面积。

符号 $\lambda$(希腊字母 Lambda)表示一个尚未确定的数,称作拉格朗日(待定)乘数。如果我们能设法确定$4x_1+2x_2=60$,约束条件得到满足,那么,无论 $\lambda$ 取何值,(12.3)最后一项都将为零。在此情况下,$Z$ 将等于 $U$。这样,避开了约束条件,对于两个变量 $x_1$ 和 $x_2$,我们就可以求 $Z$ 的自由极值来代替约束极值。问题是:我们如何才能使(12.3)中括号内的表达式为零呢?

达到此目的的办法之一就是将 $\lambda$ 视为(12.3)中的一个额外的变量,即视 $Z$ 为 $Z=Z(\lambda,x_1,x_2)$。这样,自由极值的一阶条件将由下列联立方程组:

$$\begin{aligned} Z_\lambda\ (\equiv \partial Z/\partial \lambda) &= 60-4x_1-2x_2=0 \\ Z_1(\equiv \partial Z/\partial x_1) &= x_2+2-4\lambda=0 \\ Z_2(\equiv \partial Z/\partial x_2) &= x_1-2\lambda=0 \end{aligned} \tag{12.4}$$

而且第一个方程将自动保证满足约束条件。因此,通过将约束纳入拉格朗日函数,将拉格朗日乘数作为额外变量,就可以将 $Z$ 视为三个选择变量的自由函数,通过筛选 $Z$ 的稳定值,我们便可得到约束极值 $\bar{U}$(两个选择变量)。

解(12.4),求出变量的临界值,我们得到 $\bar{x}_1=8,\bar{x}_2=14$(和 $\bar{\lambda}=4$)。正如所预料的那样,$\bar{x}_1$ 和 $\bar{x}_2$ 的值可与用代入法得到的答案进行核对。而且,显然由(12.3)可知,$\bar{Z}=128$。这与前面求得的 $\bar{U}$值是一致的。它也理应如此。

一般而言,给定目标函数

$$z=f(x,y) \tag{12.5}$$

满足约束条件

$$g(x,y)=c \tag{12.6}$$

其中 $c$ 是一个常数,[①] 我们可以将拉格朗日函数写成

$$Z = f(x,y) + \lambda [c - g(x,y)] \tag{12.7}$$

374 对于 $Z$(可把 $Z$ 视为三个变量 $\lambda$ ,$x$ 和 $y$ 的函数)的稳定值,必要条件是

$$\begin{aligned} Z_\lambda &= c - g(x,y) = 0 \\ Z_x &= f_x - \lambda g_x = 0 \\ Z_y &= f_y - \lambda g_y = 0 \end{aligned} \tag{12.8}$$

因为(12.8)中的第一个方程仅是(12.6)的重新表述,拉格朗日函数 $Z$ 的稳定值自然满足原函数 $z$ 的约束。而且因为表达式 $\lambda [c - g(x,y)]$ 现在必然为零,所以(12.7)中 $Z$ 的稳定值必然与满足约束(12.6)的(12.5)的稳定值一致。

我们用以下两个例子描述这种方法。

**例 1** 求函数

$$z = xy \qquad \text{满足约束 } x + y = 6$$

的极值。第一步是写出拉格朗日函数:

$$Z = xy + \lambda(6 - x - y)$$

为求 $Z$ 的一个稳定值,必须

$$\left.\begin{aligned} Z_\lambda &= 6 - x - y = 0 \\ Z_x &= y - \lambda = 0 \\ Z_y &= x - \lambda = 0 \end{aligned}\right\} \text{或} \left\{\begin{aligned} x + y &= 6 \\ -\lambda \quad + y &= 0 \\ -\lambda + x \quad &= 0 \end{aligned}\right.$$

---

① 还可以将常数 $c$ 归并到约束函数中,使得(12.6)表现为 $G(x,y) = 0$,其中 $G(x,y) = g(x,y) - c$。在这种情况下,(12.7)被变换成 $Z = f(x,y) + \lambda [0 - G(x,y)] = f(x,y) - \lambda G(x,y)$。选择(12.6)这种形式是因为它便于研究后面约束常数变化的比较静态效果[请参见(12.6)]。

因此,根据克莱姆法则或者其它方法,我们可求得:

$$\bar{\lambda} = 3, \qquad \bar{x} = 3, \qquad \bar{y} = 3$$

此稳定值为 $\bar{z} = \bar{z} = 9$,在我们确定它为极大值还是极小值之前,还需要对其进行二阶条件检验。这个检验我们到后面再做。

**例2** 求函数

$$z = x_1^2 + x_2^2 \qquad \text{满足 } x_1 + 4x_2 = 2$$

拉格朗日函数为

$$Z = x_1^2 + x_2^2 + \lambda\,(2 - x_1 - 4x_2)$$

稳定值的必要条件为

$$\left.\begin{aligned} Z_\lambda &= 2 - x_1 - 4x_2 = 0 \\ Z_1 &= 2x_1 - \lambda = 0 \\ Z_2 &= 2x_2 - 4\lambda = 0 \end{aligned}\right\} \text{或} \left\{\begin{aligned} x_1 + 4x_2 &= 2 \\ -\lambda + 2x_1 \qquad &= 0 \\ -4\lambda \qquad + 2x_2 &= 0 \end{aligned}\right.$$

因此由解

$$\bar{\lambda} = \frac{4}{17}, \qquad \bar{x}_1 = \frac{2}{17}, \qquad \bar{x}_2 = \frac{8}{17}$$

所定义的 $Z$ 的稳定值是 $Z = z = \frac{4}{17}$。同样,若我们要确定 $\bar{z}$ 是极大值还是极小值,还应进行二阶条件检验。

375

## 全微分法

在讨论 $z = f(x, y)$ 的自由极值时,我们知道一阶必要条件可按全微分 $dz$ 表示如下

$$dz = f_x dx + f_y dy \tag{12.9}$$

这种表述在加入约束 $g(x, y) = c$ 后仍然成立。但是在出现这个约束后,我们不能再像以前那样将 $dx$ 和 $dy$ 视作“任意”变化的

量，因为若 $g(x,y)=c$，则 $dg$ 必定等于 $dc$，因为 $c$ 是常数，$dc$ 为零。因此

$$(dg=)g_x dx+g_y dy=0 \tag{12.10}$$

而且这个联系使 $dx$ 和 $dy$ 彼此相关。这样，一阶必要条件变成 $dz=0$[(12.9)]，满足约束 $g=c$，因此也满足 $dg=0$[(12.10)]。考察 (12.9)和(12.10)，应该明确，为满足这一必要条件，我们必须有

$$\frac{f_x}{g_x}=\frac{f_y}{g_y} \tag{12.11}$$

解(12.10)求 $dy$，并将此结果代入(12.9)，可以验证这一结果。条件(12.11)与约束 $g(x,y)=c$ 一起构成了两个方程，由此可以求出 $x$ 和 $y$ 的临界值。[①]

全微分法能够与拉格朗日乘数法得出同样的一阶条件吗？我们将(12.8)与刚得到的结果相比较。(12.8)中的第一个方程仅是重复约束条件；新的结果也需满足这一约束。(12.8)中后两个方程可分别重写成

$$\frac{f_x}{g_x}=\lambda \quad 和 \quad \frac{f_y}{g_y}=\lambda \tag{12.11'}$$

且它们与(12.11)传达完全相同的信息。但要注意全微分法只能得到 $\bar{x}$ 值和 $\bar{y}$ 值，而拉格朗日乘数法不仅可以得到 $\bar{x}$ 和 $\bar{y}$ 值，还可以得到一个直接的副产品 $\bar{\lambda}$ 值。实际上 $\bar{\lambda}$ 还是 $\bar{Z}$(和 $\bar{z}$)对约束

---

① 注意，约束 $g=c$ 仍可以(12.11)一起考察，尽管我们在推导(12.11)时，我们利用了方程 $dg=0$，即(12.10)。虽然 $g=c$ 必然意味着 $dg=0$，但反之不成立：即 $dg=0$ 仅意味着 $g=$某个常数(不一定为 $c$)。除非明确地考察约束条件，否则此题就不考虑这些信息。

变化敏感性的度量，对此我们马上就将证明。因此，拉格朗日乘数法由于其解中内在地包含比较静态分析的内容，所以具有特别的优势。376

## 拉格朗日乘数的解释

为证明$\bar{\lambda}$确实度量$\bar{Z}$对约束条件变化的敏感性，我们对一阶条件(12.8)进行比较静态分析。因为$\lambda$、$x$、$y$均为内生变量，唯一合适的外生变量是约束参数$c$。参数$c$的变化将导致$xy$平面中约束曲线的移动，从而使最优解改变。特别地，$c$增加（预算增加或生产配额增加）的影响表明约束条件的放宽如何影响最优解。

要进行比较静态分析，我们仍需求助于隐函数定理。将(12.8)中三个方程取$F^j(\lambda, x, y; c)=0 (j=1,2,3)$的形式，并假定其具有连续偏导数，我们首先检验下列内生变量雅可比行列式（其中$f_{xy}=f_{yx}$及$g_{xy}=g_{yx}$）

$$|J|=\begin{vmatrix}\frac{\partial F^1}{\partial\lambda} & \frac{\partial F^1}{\partial x} & \frac{\partial F^1}{\partial y}\\ \frac{\partial F^2}{\partial\lambda} & \frac{\partial F^2}{\partial x} & \frac{\partial F^2}{\partial y}\\ \frac{\partial F^3}{\partial\lambda} & \frac{\partial F^3}{\partial x} & \frac{\partial F^3}{\partial y}\end{vmatrix}=\begin{vmatrix}0 & -g_x & -g_y\\ -g_x & f_{xx}-\lambda g_{xx} & f_{xy}-\lambda g_{xy}\\ -g_y & f_{xy}-\lambda g_{xy} & f_{yy}-\lambda g_{yy}\end{vmatrix} \tag{12.12}$$

在最优状态下不为零。当然，此时我们对此还一无所知。但我们前面已有的关于最优化问题比较静态分析的经验[参见(11.42)节的讨论]表明，这个雅可比行列式与二阶充分条件密切相关，且如果满足充分条件，则雅可比行列式在均衡（最优）状态为非零。我们把对此事实的证明留至下一节，在$|J|\neq 0$的假设下推导下面的

内容。若$|J|\neq 0$,则我们可将 $\bar{\lambda}$ ,$\bar{x}$ 和 $\bar{y}$ 表示成参数 $c$ 的隐函数:

$$\bar{\lambda}=\bar{\lambda}(c) \qquad \bar{x}=\bar{x}(c) \quad 和 \quad \bar{y}=\bar{y}(c)$$

它们均具有连续导数。我们还有恒等式 (12.13)

$$c-g(\bar{x},\bar{y})\equiv 0$$

$$f_x(\bar{x},\bar{y})-\bar{\lambda}g_x(\bar{x},\bar{y})\equiv 0 \qquad (12.14)$$

$$f_y(\bar{x},\bar{y})-\bar{\lambda}g_x(\bar{x},\bar{y})\equiv 0$$

现在,因 $Z$ 的最优值取决于$\bar{\lambda}$ ,$\bar{x}$ 和 $\bar{y}$,即

$$\bar{Z}=f(\bar{x},\bar{y})+\bar{\lambda}[c-g(\bar{x},\bar{y})] \qquad (12.15)$$

377 由(12.13),我们可将 $\bar{Z}$ 视为 $c$ 的函数。将 $\bar{Z}$ 对 $c$ 全微分,我们得到:

$$\frac{d\bar{Z}}{dc}=f_x\frac{d\bar{x}}{dc}+f_y\frac{d\bar{y}}{dc}+[c-g(\bar{x},\bar{y})]\frac{d\bar{\lambda}}{dc}+\bar{\lambda}\left(1-g_x\frac{d\bar{x}}{dc}-g_y\frac{d\bar{y}}{dc}\right)$$

$$=(f_x-\bar{\lambda}g_x)\frac{d\bar{x}}{dc}+(f_y-\bar{\lambda}g_y)\frac{d\bar{y}}{dc}+[c-g(\bar{x},\bar{y})]\frac{d\bar{\lambda}}{dc}+\bar{\lambda}$$

其中 $f_x,f_y,g_x,g_y$ 均在最优处计值。但根据(12.14),上式右边的前三项全部消失,因此我们得到一个简单的结果

$$\frac{d\bar{Z}}{dc}=\bar{\lambda} \qquad (12.16)$$

这就证实了我们的结论:拉格朗日乘数的解值是由参数 $c$ 引起的约束条件变化对目标函数最优值影响的度量。

但这里也许应当提醒一次。对于这个 $\bar{\lambda}$的解释,你特别应强调 (12.7)中的 $Z$。特别地,应把(12.7)中的最后一项写成 $\lambda[c-g(x,y)]$,而不能写成$\lambda[g(x,y)-c]$。

## $n$ 个变量和多重约束的情况

如果我们将选择变量写上下标符号,就可以很容易地将拉格

朗日乘数法推广至 $n$ 个变量的情况。这样，目标函数将是这种形式

$$z = f(x_1, x_2, \cdots, x_n)$$

满足约束条件

$$g(x_1, x_2, \cdots, x_n) = c$$

由此可知，拉格朗日函数将为

$$Z = f(x_1, x_2, \cdots, x_n) + \lambda [c - g(x_1, x_2, \cdots, x_n)]$$

其一阶条件将由下述$(n+1)$个联立方程构成：

$$Z_\lambda = c - g(x_1, x_2, \cdots, x_n) = 0$$

$$Z_1 = f_1 - \lambda g_1 = 0$$

$$Z_2 = f_2 - \lambda g_2 = 0$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$Z_n = f_n - \lambda g_n = 0$$

同样，上述方程中的第一个方程可以确保约束条件得到满足，虽然我们将把注意力集中于自由拉格朗日乘数。

当存在不止一个约束条件时，如果我们在拉格朗日函数中引入数量与约束条件相同的乘数，我们同样可以应用拉格朗日乘数 378 法。我们令 $n$ 个变量函数同时满足两个约束。

$$g(x_1, x_2, \cdots, x_n) = c \quad \text{和} \quad h(x_1, x_2, \cdots, x_n) = d$$

则用 $\lambda$ 和 $\mu$(希腊字母 mu)表示两个待定乘数，我们可构建如下拉格朗日乘数：

$$Z = f(x_1, x_2, \cdots, x_n) + \lambda [c - g(x_1, x_2, \cdots, x_n)] + \mu [d - h(x_1, x_2, \cdots, x_n)]$$

若满足两个约束条件，即拉格朗日函数中的最后两项均为零，则此

函数将与原目标函数 $f$ 有相同的值。

将 $\lambda$ 和 $\mu$ 视为两个变量，我们现在共有 $(n+2)$ 个变量，因此这种情况下的一阶条件将由 $(n+2)$ 个联立方程构成：

$$Z_\lambda = c - g(x_1, x_2, \cdots, x_n) = 0$$

$$Z_\mu = d - h(x_1, x_2, \cdots, x_n) = 0$$

$$Z_i = f_i - \lambda g_i - \mu h_i = 0 \qquad (i = 1, 2, \cdots, n)$$

这些方程可使我们解得所有的 $x_i$ 和 $\lambda$ 及 $\mu$。同前面一样，必要条件中的前两个方程实质上只是两个约束条件的重述。

## 练习 12.2

---

1 运用拉格朗日乘数法求 $z$ 的稳定值：

($a$) $z = xy$ 满足约束 $x + 2y = 2$

($b$) $z = x(y+4)$ 满足约束 $x + y = 8$

($c$) $z = x - 3y - xy$ 满足约束 $x + y = 6$

($d$) $z = 7 - y + x^2$ 满足约束 $x + y = 0$

2 在上题中，约束条件略微放松是增加还是减少了 $z$ 的最优值，增加或降低的速率是多少？

3 写出下列函数的拉格朗日函数和稳定值的一阶条件（不必解方程）：

($a$) $z = x + 2y + 3w + xy - yw$

满足约束 $x + y + 2w = 10$

($b$) $z = x^2 + 2xy + yw^2$

满足约束 $2x + y + w^2 = 24$ 和 $x + w = 8$

4 若将约束条件写成 $G(x, y) = 0$，而不写成 $g(x, y) = c$ 的形式，那么，拉格朗日函数和一阶条件应如何修正？

5 在讨论全微分法时，我们曾指出已知约束条件 $g(x, y) = c$，可以推导出 $dg = 0$。同理我们还可以进一步推导出 $d^2 g = d(dg) = d(0) = 0$。但前

面在讨论函数 $z=f(x,y)$ 的非约束极值时，我们得到这样一种情况：其中 $dz=0$ 或者是为 $d^2z$ 的正定所伴随，或者是为其负定所伴随，但却没有 $d^2z=0$。你如何解释这两种情况处理上的差别？

6 如果拉格朗日函数被写成 $Z=f(x,y)+\lambda[g(x,y)-c]$，而不是(12.7)那种形式，我们仍可以像(12.16)那样解释拉格朗日乘数吗？ 379

## 12.3 二阶条件

把拉格朗日乘数作为附加变量，使得在自由极值中使用的一阶条件，也可以应用于约束极值问题。这就诱使人们更进一步，以借用其二阶必要条件和充分条件，但并不能这样做。因为即使 $\bar{Z}$ 对选择变量是一个标准的极值问题，但对拉格朗日乘数来说却**不是**。从(12.15)我们可以具体地看到，$\bar{\lambda}$不同于 $\bar{x}$ 和 $\bar{y}$，若它被$\lambda$ 的任意其它值所取代，对 $\bar{Z}$ 都不会产生影响，因为$[c-g(\bar{x},\bar{y})]$恒为零。因此在最优解中 $\lambda$ 所充当的角色与 $x$ 和 $y$ 有根本不同。[①] 虽然在讨论一阶条件时将 $\lambda$ 视为另一个选择变量没有害处，但我们必须小心，不能盲目地把自由极值中使用的二阶条件应用于现有约束条件的情况。而且，我们必须推导出新的二阶条件。正如我们将看到的，这样的条件是二阶全微分 $d^2z$ 表述的。但由于约

① 在下一章将要讨论的被称作"非线性规划"的更一般的约束最优化的框架中，我们将要证明，在具有不等式约束时，若 $\bar{Z}$ 是对 $x$ 和 $y$ 的极大值(或极小值)，那么，它实际上是对 $\lambda$ 的极小值(极大值)。换言之，点$(\bar{\lambda},\bar{x},\bar{y})$是个鞍点。现在的情况(其中 $\bar{Z}$ 是对 $x$ 和 $y$ 的真正极值，但对$\lambda$ 却是一个变量)可以看作是鞍点的变形。鞍点解$(\bar{\lambda},\bar{x},\bar{y})$的性质也可以引出一个重要的概念——"对偶性"。但这一问题最好还是在后面探讨。

束条件的存在,判别准备需要作一些重要的修正。

## 二阶全微分

我们曾指出,由于约束 $g(x,y)=c$ 意味着 $dg=g_x dx+g_y dy=0$,如(12.10)中那样,$dx$ 和 $dy$ 不再都是任意的。当然,我们仍可以将 $dx$ 看成任意变化的,但 $dy$ 必被视为取决于 $dx$,且总可以选择 $dx$,以满足(12.10),即满足 $dy=-(g_x/g_y)dx$。从另一角度看,一旦 $dx$ 值设定,$dy$ 将取决于 $g_x$ 和 $g_y$,但因导数 $g_x$ 和 $g_y$ 又依赖于变量 $x$ 和 $y$,所以,$dy$ 也将依赖于 $x$ 和 $y$。显然前面(11.6)中 $d^2z$ 的公式是以 $dx$ 和 $dy$ 为任意值为基础的,所以它不再适用了。

为求出 $d^2z$ 的新的合适的表达式,我们必须在微分时(若 $dx$ 保持不变)将 $dy$ 视为依赖于 $x$ 和 $y$ 的变量。那么,

380

$$
\begin{aligned}
d^2z=d(dz)&=\frac{\partial(dz)}{\partial x}dx+\frac{\partial(dz)}{\partial y}dy\\
&=\frac{\partial}{\partial x}(f_x dx+f_y dy)dx+\frac{\partial}{\partial y}(f_x dx+f_y dy)dy\\
&=\left[f_{xx}dx+\left(f_{xy}dy+f_y\frac{\partial dy}{\partial x}\right)\right]dx\\
&\qquad+\left[f_{yx}dx+\left(f_{yy}dy+f_y\frac{\partial dy}{\partial y}\right)\right]dy\\
&=f_{xx}dx^2+f_{xy}dydx+f_y\frac{\partial(dy)}{\partial x}dx+f_{yx}dxdy\\
&\qquad+f_{yy}dy^2+f_y\frac{\partial(dy)}{\partial y}dy
\end{aligned}
$$

因第三项和第六项可以简化为

$$f_y\left[\frac{\partial(dy)}{\partial x}dx + \frac{\partial(dy)}{\partial y}dy\right] = f_y d(dy) = f_y d^2 y$$

所以所求的 $d^2 z$ 表达式为

$$d^2 z = f_{xx} dx^2 + 2f_{xy} dxdy + f_{yy} dy^2 + f_y d^2 y \tag{12.17}$$

它与(12.6)的不同之处仅在于最后一项，$f_y d^2 y$。

应注意的是，最后一项是一次的[$d^2 y$ 不同于$(dy)^2$]，因此，由于(12.17)中存在 $d^2 y$，所以 $d^2 z$ 不是二次型。但利用约束 $g(x,y)=c$，可将 $d^2 z$ 变换成二次型。因为约束意味着 $dg=0$ 和 $d^2 g=d(dg)=0$，所以利用获得(12.17)的步骤，可以得到

$$(d^2 g =) g_{xx} dx^2 + 2g_{xy} dxdy + g_{yy} dy^2 + g_y d^2 y = 0$$

解这个方程得到 $d^2 y$，再将此结果代入(12.17)，我们可以去掉一次表达式 $d^2 y$，并将 $d^2 z$ 写成如下二次型：

$$d^2 z = \left(f_{xx} - \frac{f_y}{g_y}g_{xx}\right)dx^2 + 2\left(f_{xy} - \frac{f_y}{g_y}g_{xy}\right)dxdy$$
$$+ \left(f_{yy} - \frac{f_y}{g_y}g_{yy}\right)dy^2$$

由于(12.11′)，第一个加括号的系数可以化简为$(f_{xx} - \lambda g_{xx})$，其它各项也可作类似化简。但对(12.8)中的导数进行偏微分，可以求出下列二阶导数：

$$\begin{aligned} Z_{xx} &= f_{xx} - \lambda g_{xx} \\ Z_{xy} &= f_{xy} - \lambda g_{xy} = Z_{yx} \\ Z_{yy} &= f_{yy} - \lambda g_{yy} \end{aligned} \tag{12.18}$$

它们恰为括号中的系数。因此，利用拉格朗日函数，我们最终可将 $d^2 z$ 更简洁地表示如下：

$$d^2z = Z_{xx}dx^2 + Z_{xy}dxdy + Z_{yx}dydx + Z_{yy}dy^2 \qquad (12.17')$$

381 (12.17′)的系数仅是 $Z$ 对选择变量 $x$ 和 $y$ 的二阶偏导数；因此，它们合在一起可产生一个海塞行列式。

## 二阶条件

对于 $z=f(x,y)$，满足约束 $g(x,y)=c$ 的约束极值，其二阶必要条件和充分条件仍由在稳定点计值的二阶全导数 $d^2z$ 的代数符号所决定。但这里有一个非常重要的变化。我们这里关注的不是对 $dx$ 和 $dy$（不同时为零）的所有可能值的 $d^2z$ 的有定性或半定性符号，而是满足线性约束(12.10)，即 $g_xdx+g_ydy=0$ 的那些 $dx$ 和 $dy$（不同时为零）值的 $d^2z$ 的符号。因此，二阶必要条件为：

对于 $z$ 的极大值：$d^2z$ 为负半定，满足 $dg=0$

对于 $z$ 的极小值：$d^2z$ 为正半定，满足 $dg=0$

二阶充分条件为：

对于 $z$ 的极大值：$d^2z$ 为负定，满足 $dg=0$

对于 $z$ 的极小值：$d^2z$ 为正定，满足 $dg=0$ 下面，我们将集中讨论二阶充分条件。

由于满足约束 $g_xdx+g_ydy=0$ 的数偶$(dx,dy)$仅构成了 $dx$ 和 $dy$ 可行集合中的一个子集，因此约束的有定性符号相对于上一章所讨论的非约束的有定性符号而言是不严格的，即更易于满足的。换言之，约束极值的二阶充分条件比自由极值的条件要弱。这是一个好消息，因为充分条件与必要条件不同，必要条件必须非常严格以使之成为一种有效的筛选机制，而充分条件应当弱一些，

从而使极值得以存在。[1]

## 海塞加边行列式

如自由极值中的情况一样，可以用行列式表示约束极值的二阶充分条件。但在约束极值情况下，我们遇到的是所谓的“海塞加边行列式”，而非海塞行列式$|H|$。 382

为发展这一思想作准备，我们首先分析满足线性约束的两个变量的二次型，比如

$$q = au^2 + 2huv + bv^2 \qquad \text{满足约束} \qquad \alpha u + \beta v = 0$$

的有定性符号的条件。因为由约束条件有 $v = -(\alpha/\beta)u$，我们可以将 $q$ 写成仅有一个变量的函数：

$$q = au^2 - 2h\frac{\alpha}{\beta}u^2 + b\frac{\alpha^2}{\beta^2}u^2 = (a\beta^2 - 2h\alpha\beta + b\alpha^2)\frac{u^2}{\beta^2}$$

很明显，当且仅当括号中的表达式为正(负)时，$q$ 为正(负)定。现在，恰巧下列对称行列式

$$\begin{vmatrix} 0 & \alpha & \beta \\ \alpha & a & h \\ \beta & h & b \end{vmatrix} = 2h\alpha\beta - a\beta^2 - b\alpha^2$$

正好是上面所提到的表达式的负数。因此，我们可以表述为

$$\text{当且仅当} \qquad \begin{vmatrix} 0 & \alpha & \beta \\ \alpha & a & h \\ \beta & h & b \end{vmatrix} \begin{cases} <0 \\ >0 \end{cases}$$

---

[1] “百万美元银行存款”对于“能够吃得起牛排晚餐”而言，显然是一个充分条件。但这个条件极其有限的适应性使之没有什么实际价值。更有意义的充分条件也许应当像“钱包里的 20 美元”，它没有那么严格的财政要求。

$$q \text{ 为} \begin{Bmatrix} \text{正定} \\ \text{负定} \end{Bmatrix} \text{满足 } \alpha u + \beta v = 0$$

值得注意的是，此判别标准中所运用的判别式只不过是由原来二次型的判别式 $\begin{vmatrix} a & h \\ h & b \end{vmatrix}$，上面和左面加上相同的边所构成。而且，所加的边由约束条件的两个系数 $\alpha$ 和 $\beta$ 构成，并在主对角线上加上一个零。此加边判别式是对称的。

**例 1** 确定

$q = 4u^2 + 4uv + 3v^2$ 满足约束 $u - 2v = 0$ 是正定还是负定。

我们首先形成加边判别式：$\begin{vmatrix} 0 & 1 & -2 \\ 1 & 4 & 2 \\ -2 & 2 & 3 \end{vmatrix}$，此行列式由在原二次型行列式中加入约束条件 $uv$ 的系数所构成。由于此行列式具有负值$(-27)$，所以 $q$ 必为正定。

当将此判别标准应用于(12.17′)中的二次型 $d^2z$ 时，变量 $u$ 和 $v$ 分别变成 $dx$ 和 $dy$，一般判别式由海塞矩阵 $\begin{vmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{vmatrix}$ 所构成。而且对二次型的约束为 $g_x dx + g_y dy = 0$，所以，有 $\alpha = g_x$，383 $\beta = g_y$。因此，对于满足上文提到的约束的 $dx$ 和 $dy$ 值，我们现在有如下关于 $d^2z$ 定性符号的行列式判别准则：

$$\text{当且仅当} \quad \begin{vmatrix} 0 & g_x & g_y \\ g_x & Z_{xx} & Z_{xy} \\ g_y & Z_{yx} & Z_{yy} \end{vmatrix} \begin{Bmatrix} <0 \\ >0 \end{Bmatrix}$$

$$d^2z \text{ 为} \begin{Bmatrix} \text{正定} \\ \text{负定} \end{Bmatrix} \text{满足 } dg = 0$$

上面的行列式，通常称作海塞加边行列式，用符号 $|\bar{H}|$ 表示，其中

$H$ 上面的横杠"-"表示边。在此基础上,我们可以得出结论:给定 $z=f(x,y)$ 或 $Z=f(x,y)+\lambda[c-g(x,y)]$ 的稳定值,$|\bar{H}|$ 为正是此稳定值为 $z$ 的极大值的充分条件;类似地,$|\bar{H}|$ 为负是此稳定值为 $z$ 的极小值的充分条件;$|\bar{H}|$ 中所包含的导数均在 $x$ 和 $y$ 的临界值计值。

现在，我们已推导出二阶充分条件。正如前面所指出的那样，验证满足此条件可确保内生变量雅可比行列式在最优状态不为零，是非常容易的事。将(12.18)代入(12.12),并以 -1 同时乘雅可比行列式的第一列和第一行(这不改变行列式的值),可以看出

$$|J| = \begin{vmatrix} 0 & g_x & g_y \\ g_x & Z_{xx} & Z_{xy} \\ g_y & Z_{yx} & Z_{yy} \end{vmatrix} = |\bar{H}| \tag{12.19}$$

即内生变量雅可比行列式与海塞加边行列式相等,这是一个与(11.42)类似的结果。与自由极值相联系的(11.42)表明内生变量雅可比行列式与一般海塞行列式是相等的。但是,若在满足充分条件时,在最优状态有 $|\bar{H}|\neq 0$,则 $|J|$ 必然也不等于零。因而,在这里应用隐函数定理,以条件 $|H|\neq 0$ 代替通常的条件 $|J|\neq 0$ 是不会出错的。下面分析约束最优化问题时,我们就将采用这种作法。

**例 2** 现在我们回到 12.2 节例 1,确定那里求出的稳定值是极大值还是极小值。因 $Z_x=y-\lambda$ ,$Z_y=x-\lambda$ ,所以二阶偏导数为 $Z_{xx}=0,Z_{xy}=Z_{yx}=1,Z_{yy}=0$。我们需要的加边元素为 $g_x=1$,$g_y=1$。因此求得

$$|\bar{H}| = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix} = 2 > 0$$

这证明 $\bar{z}=9$ 为极大值。

384 **例 3** 继续作 12.2 节中的例 2。我们看到 $Z_1=2x_1-\lambda$ 和 $Z_2=2x_2-4\lambda$ 。由此得 $Z_{11}=2, Z_{12}=Z_{21}=0, Z_{22}=2$。由约束 $x_1+4x_2=2$,可以得到 $g_1=1$ 和 $g_2=4$。由此可知海塞加边行列式为

$$|\bar{H}| = \begin{vmatrix} 0 & 1 & 4 \\ 1 & 2 & 0 \\ 4 & 0 & 2 \end{vmatrix} = -34 < 0$$

所以 $\bar{z}=\frac{4}{17}$为极小值。

### $n$ 个变量的情况

当目标函数形式为

$z=f(x_1, x_2\cdots, x_n)$,满足 $g(x_1, x_2, \cdots, x_n)=c$ 二阶条件仍取决于$d^2z$ 的符号。因为 $d^2z$ 是一个变量为$dx_1, dx_2, \cdots, dx_n$ 的约束二次型,满足关系

$$(dg=)g_1dx_1+g_2dx_2+\cdots+g_ndx_n=0$$

$d^2z$ 正定或负定的条件中仍包含海塞加边行列式。但这次,这些条件必须表示成海塞加边主子式。

给定海塞加边行列式

$$|\bar{H}| = \begin{vmatrix} 0 & g_1 & g_2 & \cdots & g_n \\ g_1 & Z_{11} & Z_{12} & \cdots & Z_{1n} \\ g_2 & Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ g_n & Z_{n1} & Z_{n2} & \cdots & Z_{nn} \end{vmatrix}$$

其逐次加边主子式定义为

$$|\bar{H}_2| \equiv \begin{vmatrix} 0 & g_1 & g_2 \\ g_1 & Z_{11} & Z_{12} \\ g_2 & Z_{21} & Z_{22} \end{vmatrix} \qquad |\bar{H}_3| \equiv \begin{vmatrix} 0 & g_1 & g_2 & g_3 \\ g_1 & Z_{11} & Z_{12} & Z_{13} \\ g_2 & Z_{21} & Z_{22} & Z_{23} \\ g_3 & Z_{31} & Z_{32} & Z_{33} \end{vmatrix} \quad (等)$$

而最后一项加边主子式为 $|\bar{H}_n| = |H|$。在新引入的符号中，$H$ 上面的符号依然表示加边，而下标表示加边主子式的阶。例如，$|\bar{H}_2|$ 包括海塞行列式二阶主子式，以及加边 $0, g_1, g_2$。其它情况也类似。$d^2z$ 为正定和负定的条件则为

$$当且仅当\begin{cases} |\bar{H}_2|, |\bar{H}_3|, \cdots, |\bar{H}_n| < 0 \\ |\bar{H}_2| > 0; |\bar{H}_3| < 0; |\bar{H}_4| > 0; 等等。\end{cases}$$

$$d^2z \text{ 为}\begin{Bmatrix} 正定 \\ 负定 \end{Bmatrix}满足\ dg = 0。$$

385

**表 12.1 相对约束极值的行列式检验**

$z = f(x_1, x_2 \cdots, x_n)$，满足 $g(x_1, x_2, \cdots, x_n) = c$；具有拉格朗日函数。

$$Z = f(x_1, x_2, \cdots, x_n) + \lambda\,[c - g(x_1, x_2, \cdots, x_n)]$$

| 条件 | 极大值 | 极小值 |
| --- | --- | --- |
| 一阶必要条件 | $Z_\lambda = Z_1 = Z_1 = \cdots = Z_n = 0$ | $Z_\lambda = Z_1 = Z_2 = \cdots = Z_n = 0$ |
| 二阶充分条件* | $\|\bar{H}_2\| > 0; \|\bar{H}_3\| < 0;$ $\|\bar{H}_4\| > 0; \cdots; (-1)^n \|\bar{H}_n\| > 0$ | $\|\bar{H}_2\|, \|\bar{H}_3\|, \cdots, \|\bar{H}_n\| < 0$ |

* 仅当一阶条件满足后方可应用。

在前一个条件中，从 $|\bar{H}_2|$ 开始的所有加边主子式必须都为负；在后一条件中，主子式的符号则必须交替改变。如前所述，$d^2z$ 为正定是 $z$ 的稳定值为极小值的充分条件，而 $d^2z$ 为负定是 $z$ 的稳定值为极大值的充分条件。

综上所述,我们把约束相对极值的条件归纳在表 12.1 中。但读者应认识到,表中所列判别标准是不完全的。因为二阶充分条件并不是必要条件,不满足那些判别标准并不能排除稳定值为极大值或极小值的可能性。然而,在许多经济应用中,这个相对不严格的二阶充分条件或者得到满足,或者假定得到满足,所以表中的条件已经足够了。读者将表 12.1 中的结果与表 11.2 中自由极值情况下的结果相比较,应该是有启发的。

## 多重约束的情况

在问题中存在一个以上的约束时,二阶条件所包含的海塞行列式也要加上一个以上的边。假设存在 $n$ 个选择变量和形式为 $g^j(x_1,\cdots,x_n)=c_j$ 的 $m$ 个约束($m<n$)。那么,拉格朗日函数将为

$$Z = f(x_1,\cdots,x_n) + \sum_{j=1}^{m}\lambda_j[c_j - g^j(x_1,\cdots,x_n)]$$

386 海塞加边行列式将表现为

$$|\bar{H}| \equiv \left|\begin{array}{cccc:cccc} 0 & 0 & \cdots & 0 & g_1^1 & g_2^1 & \cdots & g_n^1 \\ 0 & 0 & \cdots & 0 & g_1^2 & g_2^2 & \cdots & g_n^2 \\ \hdashline 0 & 0 & \cdots & 0 & g_1^m & g_2^m & \cdots & g_n^m \\ \hdashline g_1^1 & g_1^2 & \cdots & g_1^m & Z_{11} & Z_{12} & \cdots & Z_{1n} \\ g_2^1 & g_2^2 & \cdots & g_2^m & Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \hdashline g_n^1 & g_n^2 & \cdots & g_n^m & Z_{n1} & Z_{n2} & \cdots & Z_{nn} \end{array}\right|$$

其中,$g_i^j \equiv \partial g^j/\partial x_i$ 为约束函数的偏导数,带有双下标符号的 $Z$ 同以前一样,表示拉格朗日函数的二阶偏导数。注意,为清晰起见,我们已将海塞加边行列式分隔成四个区域。左上部的区域仅由零

所组成，右下部的区域只是一个普通的海塞行列式；另外两个区域则包含导数 $g_i^j$，它们对主对角线呈镜像关系，因而使整个海塞加边行列式中的元素形成对称的阵列。

从 $|\bar{H}|$ 中可以形成不同的加边主子式。同前面一样，主对角线最后一个元素为 $Z_{22}$ 的加边主子式以 $|\bar{H}_2|$ 表示。再加上一行和一列，便出现 $Z_{33}$，我们就得到 $|\bar{H}_3|$，如此等等。运用上述符号，我们可用下列 $(n-m)$ 个加边主子式的代数符号表述二阶充分条件：

$$|\bar{H}_{m+1}|, |\bar{H}_{m+2}|, \cdots, |\bar{H}_n| (= |\bar{H}|)$$

对于 $z$ 的极大值，充分条件是这些加边主子式的符号交替变换，$|\bar{H}_{m+1}|$ 的符号为 $(-1)^{m+1}$。对于 $z$ 的极小值，充分条件是这些加边主子式均取相同符号，即都为 $(-1)^m$。

注意，由于 $(-1)$ 的奇次幂与偶次幂的符号恰好相反，所以约束条件的数量是奇数还是偶数，具有重大不同。还应注意，当 $m=1$ 时，刚才所述的条件便简化为表 12.1 中所列的条件。

## 练习 12.3

---

1　利用海塞加边行列式确定练习 12.2－1 中得到的每一小题的 $z$ 的稳定值是极大值还是极小值。

2　在表述约束极大值和极小值的二阶充分条件时，我们设定了 $|\bar{H}_2|$，$|\bar{H}_3|$，$|\bar{H}_4|$ 等的代数符号，但未规定 $|\bar{H}_1|$ 的符号。写成 $|\bar{H}_1|$ 的适当表达式，并证明它总取负值。

3　回忆行列式性质 II(5.3 节)，证明：

387

(a) 适当互换 $|\bar{H}_2|$ 的两行和(或)两列，每互换一次，行列式便改变一次

符号，它可以变换成

$$\begin{vmatrix} Z_{11} & Z_{12} & g_1 \\ Z_{21} & Z_{22} & g_2 \\ g_1 & g_2 & 0 \end{vmatrix}$$

($b$) 用类似的步骤，$|\bar{H}_3|$可以变换成

$$\begin{vmatrix} Z_{11} & Z_{12} & Z_{13} & g_1 \\ Z_{21} & Z_{22} & Z_{23} & g_2 \\ Z_{31} & Z_{32} & Z_{33} & g_3 \\ g_1 & g_2 & g_3 & 0 \end{vmatrix}$$

上述结果对海塞行列式主子式的“加边”方式有何启发？

4 写出具有四个选择变量和两个约束条件的约束最优化问题的海塞加边行列式。然后分别写出 $z$ 的极大值和极小值的具体的二阶充分条件。

---

## 12.4 拟凹性与拟凸性

11.5 节表明，对于自由极值问题，知道目标函数的凹性和凸性，可以免去检验二阶条件的必要性。在约束最优化问题中，如果曲面和超曲面具有适当类型的图形，也可能免除检验二阶条件。但这次极大值所求的图形是拟凹性（而非凹性），极小值所要求的图形是拟凸性（而非凸性）。正如我们将要证明的那样，拟凹性（拟凸性）是比凹性（凸性）弱的条件。这是我们唯一的期望，因为约束极值问题所免除的二阶充分条件（仅对于那些满足 $dg=0$ 的 $dx_i$，$d^2z$ 有确定的符号），比自由极值所免除的条件（对所有 $dx_i$，$d^2z$ 有确定的符号）更弱。

### 几何特征

像凹性和凸性一样，拟凹性和拟凸性也有严格与非严格之分。

我们首先介绍这些概念的几何特征：

> 令 $u$ 和 $v$ 为函数 $f$ 定义域(凸集)中的两个不同的点，定义域中的线段 $uv$ 在函数 $f$ 图形上给出弧段 $MN$，使得点 $N$ 高于或等于点 $M$。如果弧段 $MN$ 上除点 $M$ 和 $N$ 外的所有点均高于或等于点 $M$(低于或等于点 $N$)，则称函数 $f$ 为**拟凹**(**拟凸**)函数。如果弧段 $MN$ 上除 $M$ 和 $N$ 外的所有点均严格高于点 $M$(严格低于点 $N$)，则称 $f$ 为严格拟凹(严格拟凸)函数。

388

由此定义可以明确，严格拟凹(严格拟凸)函数必为拟凹(拟凸)函数，但反之不成立。

为加深理解，我们考察图 12.3，其中各图均按一元函数绘成。在图 $a$ 中，定义域中的线段 $uv$ 在曲线上产生弧段 $MN$，使得 $N$ 高于 $M$，这个特定的弧满足严格拟凹性的条件。但是，对于符合严

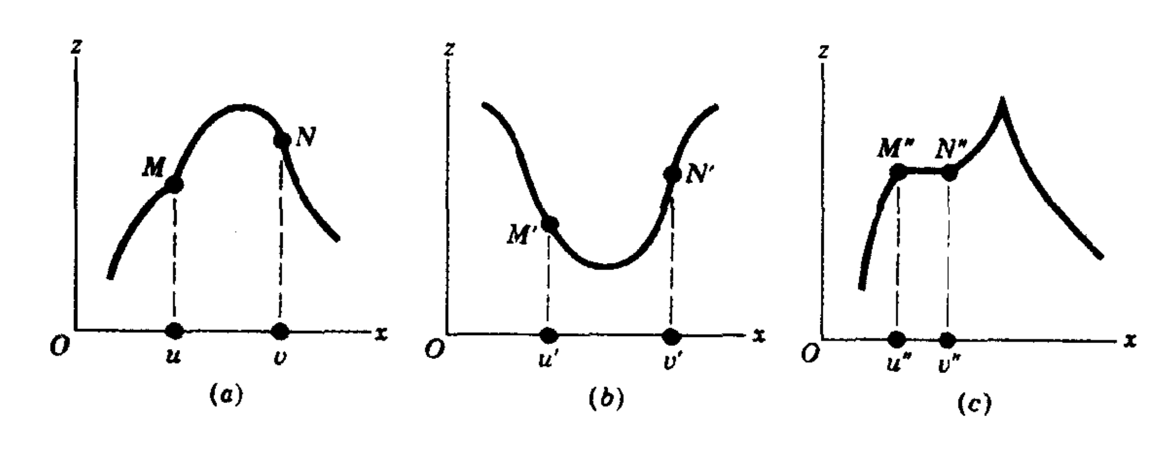



图 12.3

格拟凹条件的曲线，**所有**可能的$(u, v)$对必定具有满足同样条件的弧段。图 $a$ 中的函数确实属于这种情况。注意，此函数也满足(非严格)拟凹的条件，但它不满足拟凸性条件，因为弧 $MN$ 上的某些点高于 $N$，这是拟凸函数所不允许的。图 $b$ 中的函数具有相反的图形。弧 $M'N'$ 上的所有点均低于 $N'$ 点($N'$ 是弧 $M'N'$ 两个

端点中较高的一个点)，且所有的弧都可以绘成这种形状。因此图 $b$ 中的函数为严格拟凸函数。读者可以验证，它也满足非严格拟凸条件，但不满足拟凹性条件。使图 $c$ 有别于图 $a$ 和 $b$ 之处在于曲线中存在水平线段 $M''N''$，线段上各点高度相同。因此，线段——以及整个曲线——只满足拟凹性条件，但不满足严格拟凹条件。

概言之，拟凹但不是凹的函数的图形大致像一个钟，或者其中一部分像钟。拟凸函数的图形像一个倒置的钟，或者部分像一个倒置的钟。在钟上，同时有凹和凸的弧段是可以的(尽管并不要求这样)。这个更可以接受的特征性质使得拟凹(拟凸)性条件弱于

389

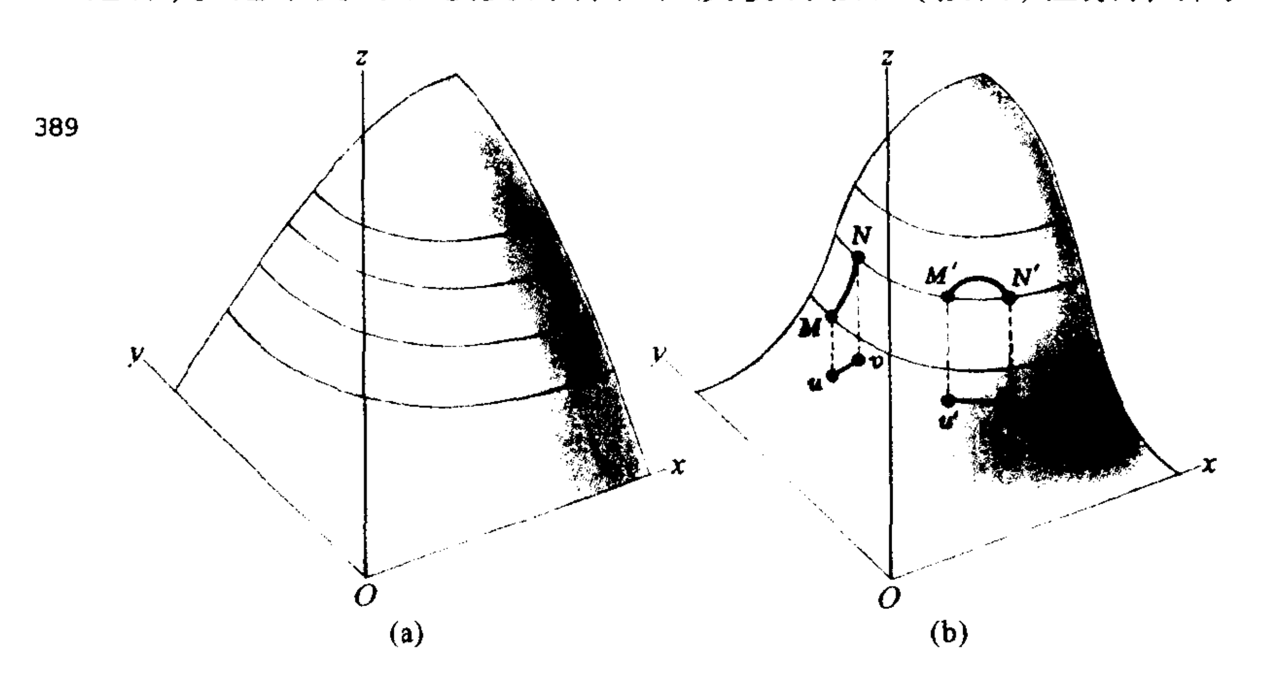



图 12.4

凹(凸)性条件。在图 12.4 中，我们在二元函数情况下将严格凹性与严格拟凹性进行了对比。如图所示，两个曲面均呈增函数，因为它们分别由山丘和钟的向上部分构成。图 $a$ 中的曲面是严格凹的，但图 $b$ 中的曲面则显然不是，因为它在接近钟的底部包含凸的部分。然而它是严格拟凹的。曲面上以 $MN$ 和 $M'N'$ 为代表的

所有弧都满足这一条件:两个端点间的每个弧上的所有点均高于较低的一个端点。回到图 $a$,我们应当注意,那里的曲面也是严格拟凹的。尽管我们没有在图 $a$ 中绘出一条说明性的弧段 $MN$ 及 $M'N'$,但不难验证所有可能的弧段都满足严格拟凹性条件。一般而言,严格凹函数必为严格拟凹函数,尽管反之不成立。我们在下面的章节中将更正式地证明这一点。

## 代数定义

上面的几何特征可以转换成更易于推广至多维情况的代数形式:

当且仅当对于函数 $f$(凸)定义域中的两个不同的点 $u$ 和 $v$,和 $0<\theta<1$,

$$f(v) \geqslant f(u) \quad \Rightarrow \quad f[\theta u + (1-\theta)v]\begin{Bmatrix} \geqslant f(u) \\ \leqslant f(v) \end{Bmatrix} \tag{12.20}$$

$f$ 为$\begin{cases} \text{拟凹} \\ \text{拟凸。} \end{cases}$

390

为使此定义适合于严格拟凹性和严格拟凸性,需将(12.20)右边两个弱不等式变成严格不等式$\begin{Bmatrix} > f(u) \\ < f(v) \end{Bmatrix}$。将(12.20)与(11.20)相比较,读者会得到有益的启示。

由此定义,下述三个定理成立。这些定理按照函数 $f(x)$,其中 $x$ 可以解释成一个向量,$x=(x_1,\cdots,x_n)$。

**定理 I (函数的负数)** 若 $f(x)$ 为拟凹(严格拟凹),则 $-f(x)$ 为拟凸(严格拟凸)。

**定理 II（凹性与拟凹性）** 任意凹（凸）函数是拟凹（拟凸）函数，但反之不成立。类似地，任意严格凹（严格凸）函数是严格拟凹（严格凸）函数，但反之亦不成立。

**定理 III（线性函数）** 若 $f(x)$是线性函数，则它既是拟凹函数，也是拟凸函数。

定理 I 以这一事实为基础：以 $-1$ 乘以一个不等式，将使不等号改变方向。令 $f(x)$ 为拟凹函数，且 $f(v) \geqslant f(u)$，则由(12.20)，有 $f[\theta u+(1-\theta)v] \geqslant f(u)$。但是，就涉及的函数 $-f(x)$而言，我们（在以 $-1$ 通乘两个不等式后）有：$-f(u) \geqslant -f(v)$和$-f[\theta u+(1-\theta)v] \leqslant -f(u)$。将 $-f(u)$看作点 $N$ 的高度，$-f(v)$看作点 $M$ 的高度，我们看到函数 $-f(x)$满足(12.20)的拟凸条件。这证明了定理 I 中四种情况之一种；其余三种情况的证明也类似。

对于定理 II，我们仅证明凹性意味着拟凹性的情况。令$f(x)$为凹函数，则由(11.20)，

$$f[\theta u+(1-\theta)v] \geqslant \theta f(u)+(1-\theta)f(v)$$

现在假设 $f(v) \geqslant f(u)$，则 $f(v)$和 $f(u)$的任意加权平均都不小于 $f(u)$，即

$$\theta f(u)+(1-\theta)f(v) \geqslant f(u)$$

综合上面两个结果，求得

$$f[\theta u+(1-\theta)v] \geqslant f(u) \quad \text{因} \quad f(v) \geqslant f(u)$$

它满足(12.20)中的拟凹定义。但要注意，该拟凹性条件不能保证凹性。

一旦定理 II 成立，便可立即得到定理 III。我们已知道，线性函数既是凹函数，也是凸函数，虽然不是严格凹和凸。根据定理

II,线性函数必然也既是拟凹,又是拟凸的,尽管不是严格拟凹和拟凸。

在凹函数和凸函数情况下,存在一个有意义的定理:即凹函数 391 (凸函数)的和也是凹(凸)函数。遗憾的是,此定理不能推广至拟凹和拟凸函数的情况。比如,两个拟凹函数的和并**不必然**是拟凹函数(参见练习 12.4-3)。

有时按如下另一种定义,也许可以更容易检验函数的拟凹性和拟凸性:

对于任意常数 $k$,当且仅当

$$\left\{\begin{array}{l} S^{\geqslant} \equiv \{x \mid f(x) \geqslant k\} \\ S^{\leqslant} \equiv \{x \mid f(x) \leqslant k\} \end{array}\right\} \tag{12.21}$$

是凸集时,函数 $f(x)$(其中 $x$ 是一个变量向量)是$\left\{\begin{array}{l}\text{拟凹的}\\ \text{拟凸的}\end{array}\right.$。

前面引入集合 $S^{\geqslant}$ 和 $S^{\leqslant}$(图 11.10)是为证明凸函数(甚至凹函数)能产生一个凸集。这里我们运用这两个集合作为检验拟凹性和拟凸性的方法。图 12.5 中的三个函数均包含凹和凸的弧段,因此它们既非凹函数,亦非凸函数。但图 $a$ 中的函数是拟凹的,因为对 $k$ 的任意值(其中仅有一个被描述),集合 $S^{\geqslant}$ 是凸集。图 $b$ 中的函数是一个拟凸函数,因为集合 $S^{\leqslant}$ 是凸集。图 $c$ 中的函数是一个单调函数,它与另两个函数的不同之处在于 $S^{\geqslant}$ 和 $S^{\leqslant}$ 均为凸集。因此,此函数既是拟凹函数,又是拟凸函数。

注意,尽管(12.21)可用于检验拟凹性和拟凸性,但不能用于区别那些性质的严格与非严格变化。还需注意,(12.21)的定义性质本身不是凹性与凸性的充分条件。特别地,给定一个凹函数,它 392 必然是拟凹函数,我们可以得出结论:$S^{\geqslant}$ 是一个凸集;但给定 $S^{\geqslant}$

是凸集，我们仅能得出结论：函数 $f$ 是拟凹（但并不必然是凹的）的。

**例1** 检验 $z=x^2(x\geqslant 0)$ 的拟凹性和拟凸性。从几何上很容易验证此函数为凸函数，而且实际上是严格凸函数。因而它是拟凸函数。有趣的是，它还是拟凹函数。因为其图形——自原点起以递增的速率增长的 $U$ 形曲线的右半部分——类似于图 12.5$c$ 能够生成凸集 $S^{\geqslant}$ 以及凸集 $S^{\leqslant}$。

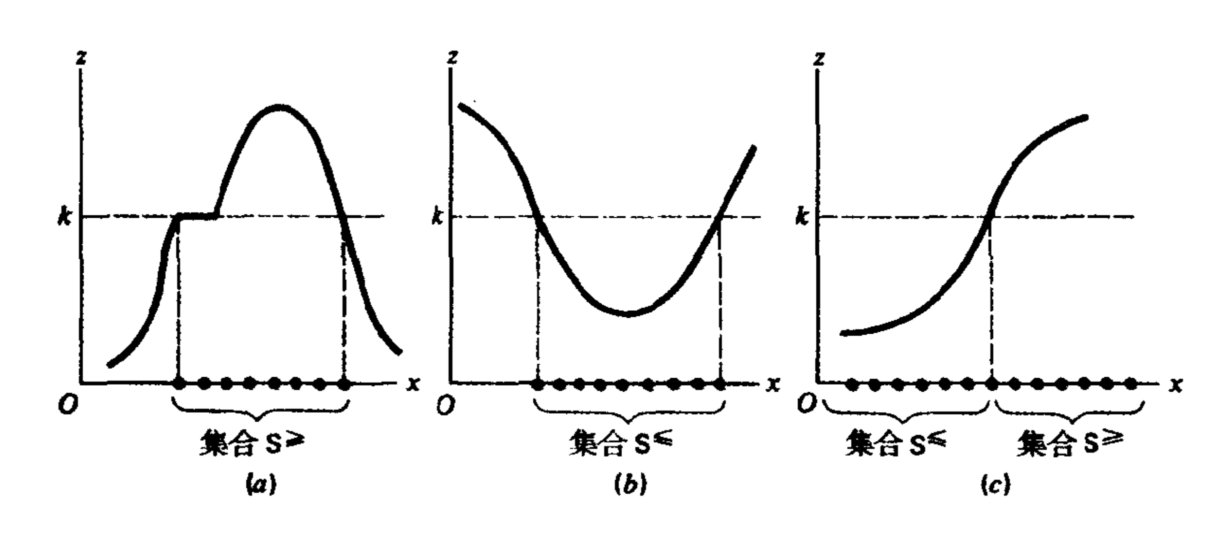



图 12.5

若我们想应用(12.20)，先令 $u$ 和 $v$ 为 $x$ 的任意两个不同的非零值，则

$$f(u)=u^2 \quad f(v)=v^2$$

和
$$f[\theta u+(1-\theta)v]=[\theta u+(1-\theta)v]^2$$

假设 $f(v)\geqslant f(u)$，即 $v^2\geqslant u^2$，则 $v\geqslant u$，或更具体地，$v>u$（因为 $u$ 和 $v$ 为不同的值）。由于加权平均值 $[\theta u+(1-\theta)v]$ 必定位于 $u$ 与 $v$ 之间，我们可以写出连续不等式

$$v^2>[\theta u+(1-\theta)v]^2>u^2 \qquad 因 \quad 0<\theta<1$$

或 $$f(v)>f[\theta u+(1-\theta)v]>f(u) \qquad 因 \quad 0<\theta<1$$

由(12.20),此结果使函数 $f$ 既是拟凹的,也是拟凸的——而且是严格拟凹和拟凸的。

**例2** 证明 $z=f(x,y)=xy \quad (x,y\geqslant 0)$ 是拟凹函数。我们将运用判别标准(12.21),并确定集合 $S^{\geqslant}\{(x,y)\mid xy\geqslant k\}$ 对任意 $k$ 值是凸集。为此,对于每个 $k$ 值,我们令 $xy=k$ 以得到一条等值曲线。同 $x$ 和 $y$ 一样,$k$ 应为非负值。在 $k>0$ 情况下,此等值曲线是 $xy$ 平面第一象限中的等轴双曲线。由等轴双曲线上及其上部的所有点组成的集合 $S^{\geqslant}$ 是个凸集。在 $k=0$ 的另一种情况下,由 $xy=0$ 定义的等值曲线是一个 $L$ 形线,且 $L$ 与 $x$ 和 $y$ 轴的非负部分重合。这时集合 $S^{\geqslant}$,包括了整个非负象限,仍是一个凸集。因此,根据(12.21),函数 $z=xy \ (x,y\geqslant 0)$ 是拟凹函数。

读者应小心,以免将等值曲线 $xy=k$(定义在 $xy$ 平面)形状与曲面 $z=xy$(定义在 $xyz$ 空间)的形状相混淆。$z$ 曲面的特征(在三维空间中为拟凹)是我们期望加以确定的,而我们对等值曲线的形状(在二维空间中对正的 $k$ 值为凸)感兴趣的只是借助于它来描述集合 $S^{\geqslant}$,以应用(12.21)中的判别标准。

**例3** 证明 $z=f(x,y)=(x-a)^2-(y-b)^2$ 为拟凸函数。令 $(x-a)^2+(y-b)^2=k$,我们知道 $k$ 必定为非负。对于每一 $k$ 值,等值曲线是 $xy$ 平面上以 $(a,b)$ 为圆心,以 $\sqrt{k}$ 为半径的圆。因为 $S^{\leqslant}=\{(x,y)\mid (x-a)^2+(y-b)^2\leqslant k\}$ 是圆及圆内所有点的集合,所以它构成了一个凸集。甚至当 $k=0$,圆退化成一个点 $(a,$ 393 $b)$ 时,也是如此,因为习惯上一个点被视作凸集。因此给定函数是拟凸函数。

## 可微函数

定义(12.20)和(12.21)并不要求函数 $f$ 的可微性。但若 $f$

可微，拟凹性和拟凸性还可以按照其一阶导数来定义：

对于一元可微函数 $f$ 定义域中的任意两个不同点 $u$ 和 $v$，当且仅当

$$f(v) \geqslant f(u) \quad \Rightarrow \quad \left\{\begin{matrix} f'(u)(v-u) \\ f'(v)(v-u) \end{matrix}\right\} \geqslant 0 \tag{12.22}$$

一元可微函数 $f(x)$ 是$\left\{\begin{matrix}\text{拟凹} \\ \text{拟凸}\end{matrix}\right\}$函数。

当右边的弱不等号变成严格不等号 $>0$ 时，定义便成为严格拟凹性与严格拟凸性的定义。当存在两个及两个以上自变量时，定义修正如下：

对于可微函数 $f(x_1,\cdots,x_n)$ 定义域中任意两个不同点 $u=(u_1,\cdots,u_n)$ 和 $v=(v_1,..,v_n)$，当且仅当

$$f(v) \geqslant f(u) \quad \Rightarrow \quad \left\{\begin{matrix} \sum_{j=1}^{n} f_j(u)(v_j-u_j) \\ \sum_{j=1}^{n} f_j(v)(v_j-u_j) \end{matrix}\right\} \geqslant 0 \tag{12.22'}$$

$f(x_1,\cdots,x_n)$ 为$\left\{\begin{matrix}\text{拟凹} \\ \text{拟凸}\end{matrix}\right\}$函数。在(12.22′)中，$f_j \equiv \partial f / \partial x_j$，在 $u$ 或 $v$ 处计值；具体在哪点计值，视情况而定。

同样，对于严格凹性和严格凸性，右边的弱不等号应改变为严格不等号 $>0$。

最后，若函数 $z=f(x_1,\cdots,x_n)$ 为二阶连续可微，拟凹性和拟凸性可以用函数的一阶导数和二阶导数（整理成加边行列式）的方

法来检验：

$$|B|=\begin{vmatrix} 0 & f_1 & f_2 & \cdots & f_n \\ f_1 & f_{11} & f_{12} & \cdots & f_{1n} \\ f_2 & f_{21} & f_{22} & \cdots & f_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ f_n & f_{n1} & f_{n2} & \cdots & f_{nn} \end{vmatrix} \qquad (12.23)$$

此加边行列式类似于上一节中介绍的海塞加边行列式$|\bar{H}|$。但 394 $|B|$与$|\bar{H}|$的不同之处在于，它的边由函数 $f$ 的一阶导数组成，而不是由外部约束函数 $g$ 组成。正是由于$|B|$仅依靠 $f$ 的一阶导数本身，所以我们可以运用$|B|$及其逐次主子式

$$|B_1|=\begin{vmatrix} 0 & f_1 \\ f_1 & f_{11} \end{vmatrix}$$

$$|B_2|=\begin{vmatrix} 0 & f_1 & f_2 \\ f_1 & f_{11} & f_{12} \\ f_2 & f_{21} & f_{22} \end{vmatrix} \quad \cdots \quad |B_n|=|B| \qquad (12.24)$$

来描述该函数构型的特征。

我们这里将表述两个条件：一个是必要条件，另一个是充分条件。两个条件均与仅由非负正交分划体(非负象限的 $n$ 维模拟)即有 $x_1,\cdots,x_n\geqslant 0$ 组成的定义域中的拟凹性和拟凸性相联系。[①]

---

① 虽然函数在凸定义域中的凹(凸)性总可以拓展至整个空间，但拟凹性和拟凸性却不能。例如，如果允许上面例 1 和例 2 中的变量取负值，则其结论便不成立。这里给出的两个条件基于 K.阿罗和 A.艾索文的文章“拟凹规划”，《经济计量学》，1961 年 10 月号 P.797，(定理 5)。他们的讨论限定于拟凹函数，我们利用拟凹函数的负值为拟凸函数这一事实，将其拓展至拟凸函数。

在非负正交分划体中，$z=f(x_1,\cdots,x_n)$要为拟凹函数，其必要条件为

$$|B_1|\leqslant 0,\quad |B_2|\geqslant 0,\quad \cdots,\quad |B_n|\begin{Bmatrix}\leqslant\\ \geqslant\end{Bmatrix}0$$

$$\text{当 } n \text{ 为}\begin{cases}\text{奇数}\\ \text{偶数}\end{cases}\qquad (12.25)$$

在非负正交分划体无论偏导数的何处计值。对于拟凸性，其必要条件为

$$|B_1|\leqslant 0,\qquad |B_2|\leqslant 0,\quad \cdots,\quad |B_n|\leqslant 0$$

$$(12.25')$$

在非负 $f$ 为拟凹的充分条件为

$$|B_1|<0,\qquad |B_2|>0,\quad \cdots,\quad |B_n|\begin{Bmatrix}<\\ >\end{Bmatrix}0 \text{ 当 } n \text{ 为}\begin{cases}\text{奇数}\\ \text{偶数}\end{cases}$$

$$(12.26)$$

在非负正交分划体无论偏导数的何处计值。对于拟凸性，相应的充分条件为

$$|B_1|<0,\qquad |B_2|<0,\quad \cdots,\quad |B_n|<0$$

$$(12.26')$$

注意，在(12.25)和(12.25′)中的条件$|B_1|\leqslant 0$是自动满足的，因为$|B_1|=-f_1^2$；这里列出它出于对称性的考虑。相反，(12.26)和(12.26′)中的条件$|B_1|<0$是不能自动满足的。

395

**例4** 函数 $z=f(x_1,x_2)=x_1x_2(x_1,x_2\geqslant 0)$是拟凹的(参见例2)。我们现在用(12.22′)对此进行检验。令 $u=(u_1,u_2)$和 $v=(v_1,v_2)$是定义域中的任意两点，则 $f(u)=u_1u_2$ 和 $f(v)=$

$v_1 v_2$。假定

$$f(v) \geqslant f(u) \quad \text{或} \quad v_1 v_2 \geqslant u_1 u_2 \quad (v_1, v_2, u_1, u_2 \geqslant 0) \tag{12.27}$$

因为 $f$ 的偏导数为 $f_1 = x_2, f_2 = x_1$,(12.22′)相当于条件

$$\begin{aligned} & f_1(u)(v_1 - u_1) + f_2(u)(v_2 - u_2) \\ = & u_2(v_1 - u_1) + u_1(v_2 - u_2) \geqslant 0 \end{aligned}$$

或者重排后

$$u_2(v_1 - u_1) \geqslant u_1(u_2 - v_2) \tag{12.28}$$

我们需要考察 $u_1$ 和 $u_2$ 值的四种可能性。首先,若 $u_1 = u_2 = 0$,则(12.28)一般是满足的。其次,若 $u_1 = 0, u_2 > 0$,则(12.28)化简为条件 $u_2 v_1 \geqslant 0$,因为 $u_2$ 和 $v_1$ 均为非负,所以,条件(12.28)依然满足。第三,若 $u_1 > 0, u_2 = 0$,则(12.28)化简为条件 $0 \geqslant -u_1 v_2$,条件仍然满足。最后,假定 $u_1$ 和 $u_2$ 均为正,从而 $v_1$ 和 $v_2$ 也为正。在(12.27)两边同时减去 $v_2 u_1$,得到

$$v_2(v_1 - u_1) \geqslant u_1(u_2 - v_2) \tag{12.29}$$

它本身又有三种子可能性:

1. 若 $u_2 = v_2$,则 $v_1 \geqslant u_1$。实际上我们应当有 $v_1 > u_1$,因为 $(u_1, u_2)$ 和 $(v_1, v_2)$ 是不同的点。这样 $u_2 = v_2$ 和 $v_1 > u_1$ 意味着条件(12.28)满足。

2. 若 $u_2 > v_2$,则由(12.29),必然也有 $v_1 > u_1$,以 $u_2/v_2$ 同乘(12.29)两边,得到

$$\begin{aligned} u_2(v_1 - u_1) & \geqslant \frac{u_2}{v_2} u_1(u_2 - v_2) \\ & > u_1(u_2 - v_2) \qquad \left[\text{因为} \frac{u_2}{v_2} > 1\right] \end{aligned} \tag{12.30}$$

因此(12.28)又得到满足。

3. 最后一种子可能性是 $u_2 < v_2$，这意味着 $u_2/v_2$ 是一正分数。在此情况下，(12.30)的第一行成立。第二行也成立，但原因不同：负数$(u_2 - v_2)$的分数$(u_2/v_2)$大于$(u_2 - v_2)$自身。

由于(12.28)在可能发生的各种情况下均得到满足，所以函数 396 $z = x_1x_2(x_1, x_2 \geqslant 0)$是拟凹的。因此，必要条件(12.25)应成立。由于 $f$ 的偏导数为

$$f_1 = x_2 \quad f_2 = x_1 \quad f_{11} = f_{22} = 0 \quad f_{12} = f_{21} = 1$$

可知相关的主子式是

$$|B_1| = \begin{vmatrix} 0 & x_2 \\ x_2 & 0 \end{vmatrix} = -x_2^2 \leqslant 0$$

$$|B_2| = \begin{vmatrix} 0 & x_2 & x_1 \\ x_2 & 0 & 1 \\ x_1 & 1 & 0 \end{vmatrix} = 2x_1x_2 \geqslant 0$$

所以(12.25)的确满足。但要注意，仅在正的正交分划体上，(12.26)才能得到满足。

**例5** 证明 $z = f(x, y) = x^a y^b (x, y > 0, 0 < a, b < 1)$是拟凹函数。此函数的偏导数为

$$f_x = ax^{a-1}y^b \quad f_y = bx^a y^{b-1} \quad f_{yy} = b(b-1)x^a y^{b-2}$$

$$f_{xx} = a(a-1)x^{a-2}y^b \quad f_{xy} = f_{yx} = abx^{a-1}y^{b-1}$$

因此，$|B|$的主子式有如下符号

$$|B_1| = \begin{vmatrix} 0 & f_x \\ f_x & f_{xx} \end{vmatrix} = -(ax^{a-1}y^b)^2 < 0$$

$$|B_2| = \begin{vmatrix} 0 & f_x & f_y \\ f_x & f_{xx} & f_{xy} \\ f_y & f_{yx} & f_{yy} \end{vmatrix} = [2a^2b^2 - a(a-1)b^2$$

$$- a^2b(b-1)]x^{3a-2}y^{3b-2} > 0$$

这满足拟凹性的充分条件(12.26)。尽管判别标准(12.26)不能证实这一点,但给定函数实际上是严格拟凹函数。

## 对海塞加边行列式的进一步考察

(12.23)定义的加边行列式$|B|$与海塞加边行列式

$$|\bar{H}| = \begin{vmatrix} 0 & g_1 & g_2 & \cdots & g_n \\ g_1 & Z_{11} & Z_{12} & \cdots & Z_{1n} \\ g_2 & Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ g_n & Z_{n1} & Z_{n2} & \cdots & Z_{nn} \end{vmatrix}$$

的不同之处有两方面:(1) $|B|$ 中的加边元素是函数 $f$ 而非 $g$ 的一阶偏导数;且 (2) $|B|$中的其余元素是 $f$ 而非拉格朗日函数 $Z$ 397 的二阶偏导数。然而,在线性约束方程 $g(x_1,\cdots,x_n) = a_1x_1 + \cdots + a_nx_n = c$ 的特定情况下[一种在经济学中经常遇到的情况(见12.5 和 12.7 节)],$Z_{ij}$简化为$f_{ij}$。这样,拉格朗日函数为

$$Z = f(x_1,\cdots,x_n) + \lambda[c - a_1x_1 - \cdots - a_nx_n]$$

所以 $$Z_j = f_j - \lambda a_j \quad 和 \quad Z_{ij} = f_{ij}$$

回到"边",我们注意到线性约束函数产生一阶导数 $g_j = a_j$。进而,当一阶条件满足,我们有 $Z_j = f_j - \lambda a_j = 0$,所以 $f_j = \lambda a_j$,或

$f_j=\lambda g_j$。因此，$|B|$中的边只不过是$|\bar{H}|$的边被正的标量$\lambda$乘。通过顺序提取$|\bar{H}|$的横边和纵边的公因子$\lambda$(见 5.3 节例 5)，得到

$$|B| = \lambda^2 |\bar{H}|$$

结果，在线性约束情况下，在 $Z$ 的稳定点，两个行列式总有相同的符号。同理，在该点主子式$|B_i|$和$|\bar{H}_i|$($i=1,\cdots,n$)必然也有相同的符号。那么由此可知，如果加边行列式$|B|$满足(12.26)中拟凹性的充分条件，则海塞加边行列式$|\bar{H}|$必定满足表 12.1 约束最大化的二阶充分条件。类似的联系还存在于拟凸性与满足线性约束最小化的二阶条件之间。

## 绝对与相对极值

拟凹性与二阶条件之间关系的更全面描述列在图 12.6 中。(对其作适当修正，它也适于拟凸性的图形。)图 12.6 与图 11.5 按照相同的思想构造起来，并按相同的方式阅读，它将拟凹性与约束绝对极大值及约束相对极大值联系起来。图上部的三个椭圆概括了相对约束极大值的一阶和二阶条件。中间那列矩形，同图 11.5 中的那些矩形一样，将相对极大值、绝对极大值、唯一绝对极大值的概念彼此联系起来。

但真正有价值的信息可以在两个菱形和通过菱形的加长的⇒符号中发现。左边的图形表明，一旦一阶条件满足，且如果菱形中所列的两个附加条件得到满足，便得到绝对约束极大值的充分条件。第一个附加条件是函数 $f$ 为显拟凹函数，这是一个新的、我们必须马上加以定义的概念。

当且仅当

$$f(v) > f(u) \quad \Rightarrow \quad f[\theta u + (1-\theta)v] > f(u)$$

398

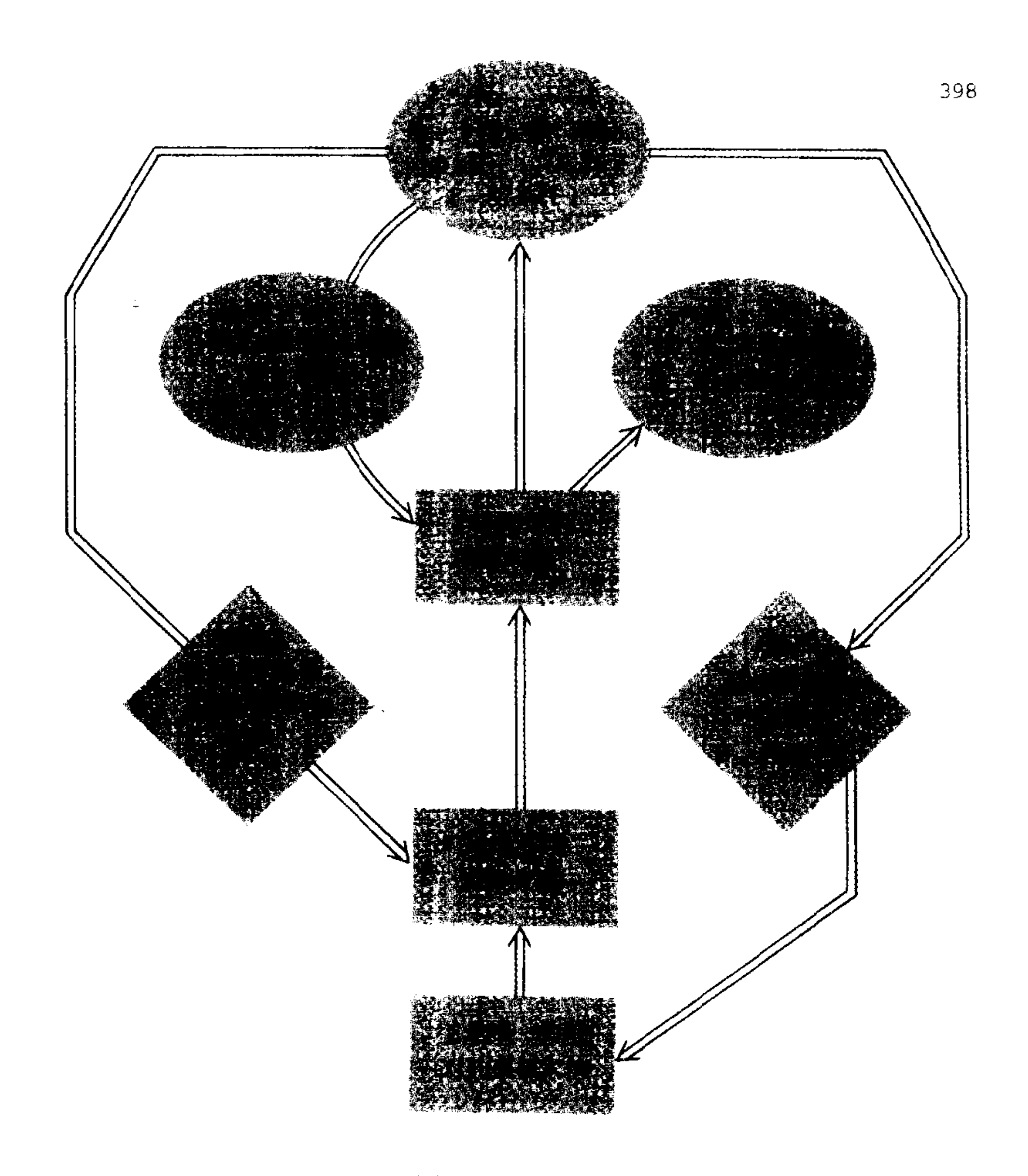


图 12.6

$f$ 为**显拟凹函数**。

这个定义性质表明，只要曲面上的点 $f(v)$ 高于另一个点 $f(u)$，那么，所有中间的点（位于定义域中线段 $uv$ 正上方曲面上的点）[399] 必定也高于 $f(u)$。这个规定的作用在于排除曲面上的任何水平

平面部分。[1] 注意，显拟凹性条件并不像严格拟凹性条件那样强，因为后者要求 $f[\theta u+(1-\theta v)]>f(u)$，即便对 $f(v)=f(u)$ 也是如此，这意味着非水平平面也被排除了。[2] 左边菱形的附加条件是集合 $\{(x_1,\cdots,x_n)\mid g(x_1,\cdots x_n)=c\}$ 为凸集。但是如果像我们目前内容中那样，把约束条件设定为等式，那么当且仅当约束函数是线性的（比如二维定义域中的一条直线）时，所说的集合才能是凸集。因此，在这里，第二个附加条件只是指线性约束方程。当这两个附加条件都满足时，我们就将研究位于定义域中直线（或平面或超平面）正上方不含水平区域的曲面（或超曲面）。在这个曲面的子集中求得的局部极大值必定为绝对约束极大值。

图 12.6 右边的菱形包含比严格拟凹性更强的条件。严格拟凹函数必定为显拟凹函数，尽管反之不成立。因此，当严格拟凹性代替显拟凹性时，仍能保证存在绝对约束极值。但这时绝对约束极值必定也是唯一的，因为曲面上任何地方均无平面部分存在从根本上排除了多重约束极大值的可能性。

## 练习 12.4

1 绘出有如下特征的严格拟凹曲线 $z=f(x)$：

(a) 也是拟凸的 (d) 不是凹的

---

① 令曲面包含水平平面部分 $P$ 使得 $f(u)\in P$，$f(v)\in P$。那么这些位于 $P$ 上的中间的点的高度将等于 $f(u)$，因而违背严格拟凹性的要求。

② 令曲面包含斜平面部分 $P'$，使得 $f(u)=f(v)$ 均位于 $P'$ 上。那么，所有中间的点也将位于 $P'$ 上，且高度等于 $f(u)$，因而违背严格拟凹性的要求。

(*b*) 不是拟凸的　　　　(*e*) 既非凹也非凸的

(*c*) 不是凸的　　　　(*f*) 既凹又凸的

2　下面的函数是拟凹的吗？是严格拟凹的吗？先用图形进行检验，再用(12.20)进行代数检验。假设 $x \geqslant 0$。

(*a*) $f(x)=a$　(*b*) $f(x)=a+bx(b>0)$　(*c*) $f(x)=a+cx^2(c<0)$

3　(*a*) 令 $z=f(x)$可绘成斜率为负，形状像第一象限内钟形曲线右半部分的那样一条曲线，并通过点(0,5)，(2,4)，(3,2)，(5,1)。令 $z=g(x)$斜率为正的45°线。$f(x)$和 $g(x)$为拟凹的吗？

(*b*) 现在绘出和 $f(x)+g(x)$的图形。此种函数是拟凹函数吗？

4　通过考察图形并运用(12.21)，检验下列函数是如下情况中的哪一种：拟凹；拟凸；既是拟凹，又是拟凸；既非拟凹，也非拟凸。

(*a*) $f(x)=x^3-2x$　(*b*) $f(x_1,x_2)=6x_1-9x_2$

(*c*) $f(x_1,x_2)=x_2-\ln x_1$

5　(*a*) 验证三次函数 $z=ax^3+bx^2+cx+d$ 一般情况下既非拟凹函数，也非拟凸函数。　400

(*b*) 对于 $x \geqslant 0$，有可能对参数施加某些限制，使得函数变成既是拟凹，又是拟凸吗？

6　运用(12.22)检验 $z=x^2(x \geqslant 0)$的拟凹性和拟凸性。

7　证明 $z=xy$ $(x,y \geqslant 0)$不是拟凹的。

8　运用加边行列式检验下列函数的拟凹性和拟凸性：

(*a*) $z=-x^2-y^2(x,y>0)$　(*b*) $z=-(x+1)^2-(y+2)^2$　$(x,y>0)$

---

## 12.5　效用最大化与消费需求

我们前面曾引用效用函数的最大化作为约束最优化的例子。现在我们来更详尽地讨论这一问题。为简单计，我们仍假设消费者仅选择两种物品，它们都有连续的正的边际效用函数。两种物品的价格均由市场决定，因而是外生的，尽管我们在本节将省略价

格符号的零下标。若消费者的购买力为给定数量 $B$,那么要解决的问题是最大化平滑(指数)效用函数

$$U = U(x, y) \qquad (U_x, U_y > 0)$$

满足

$$xP_x + yP_y = B$$

## 一阶条件

最优化模型的拉格朗日函数是

$$Z = U(x, y) + \lambda (B - xP_x - yP_y)$$

作为一阶条件,我们得到如下一组联立方程

$$\begin{aligned} Z_\lambda &= B - xP_x - yP_x = 0 \\ Z_x &= U_x - \lambda P_x = 0 \\ Z_y &= U_y - \lambda P_y = 0 \end{aligned} \tag{12.31}$$

因最后两个方程等于

$$\frac{U_x}{P_x} = \frac{U_y}{P_y} = \lambda \tag{12.31'}$$

所以一阶条件实际上要求满足(12.31′),满足预算约束——(12.31)中的第一个方程。(12.31′)所表述的不过是在古典消费 [401] 者理论中所熟悉的命题:为使效用最大化,消费者必须分配其预算以使每一物品的边际效用与价格之比率相等。具体地说,在均衡或最优状态,这些比率应有共同值 $\bar{\lambda}$。正如我们前面所知道的那样,$\bar{\lambda}$度量约束常数对目标函数最优值的比较静态效应。因此,在这里 $\bar{\lambda} = (\partial \bar{U} / \partial B)$;即拉格朗日乘数的最优值可以解释成当消费者效用最大化时,货币(预算货币)的边际效用。

若我们将(12.31′)中的条件重述如下

$$\frac{U_x}{U_y} = \frac{P_x}{P_y} \qquad (12.31'')$$

则按照无差异曲线的概念,我们可以对一阶条件进行新的解释。

**无差异曲线**是指能够产生相同效用水平 $U$ 的 $x$ 与 $y$ 组合的点的轨迹。这意味着在一条无差异曲线上必然有

$$dU = U_x dx + U_y dy = 0$$

它隐含着 $dy/dx = -U_x/U_y$。因此,若我们在 $xy$ 平面上绘出一条无差异曲线,那么,其斜率 $dy/dx$,必定等于边际效用比率的负值。(因为我们假定 $U_x, U_y > 0$,无差异曲线的斜率必定为负。)反之,因为 $U_x/U_y$ 是无差异曲线斜率的负值,所以它必然表示两物品间的**边际替代率**。

$P_x/P_y$ 是何含义呢?正如我们将要看到的那样,此比率表示预算约束图形斜率的负值。预算约束 $xP_x + yP_y = B$,还可以写成 402

$$y = \frac{B}{P_y} - \frac{P_x}{P_y}x$$

所以当像图 12.7 中那样在 $xy$ 平面上绘出预算线时,它是一条斜率为 $-P_x/P_y$,纵截距为 $B/P_y$ 的直线。

因此,一阶条件的新形式——(12.31″)加上预算约束——揭示出要使效用最大化,消费者必须对其预算进行分配,以使预算线的斜率(消费者必须将其支出控制在预算线上)等于无差异曲线的斜率。图 12.7$a$ 中的 $E$ 点,即预算线与无差异曲线的切点,满足这一条件。

## 二阶条件

若上述问题的海塞加边行列式为正,即如果

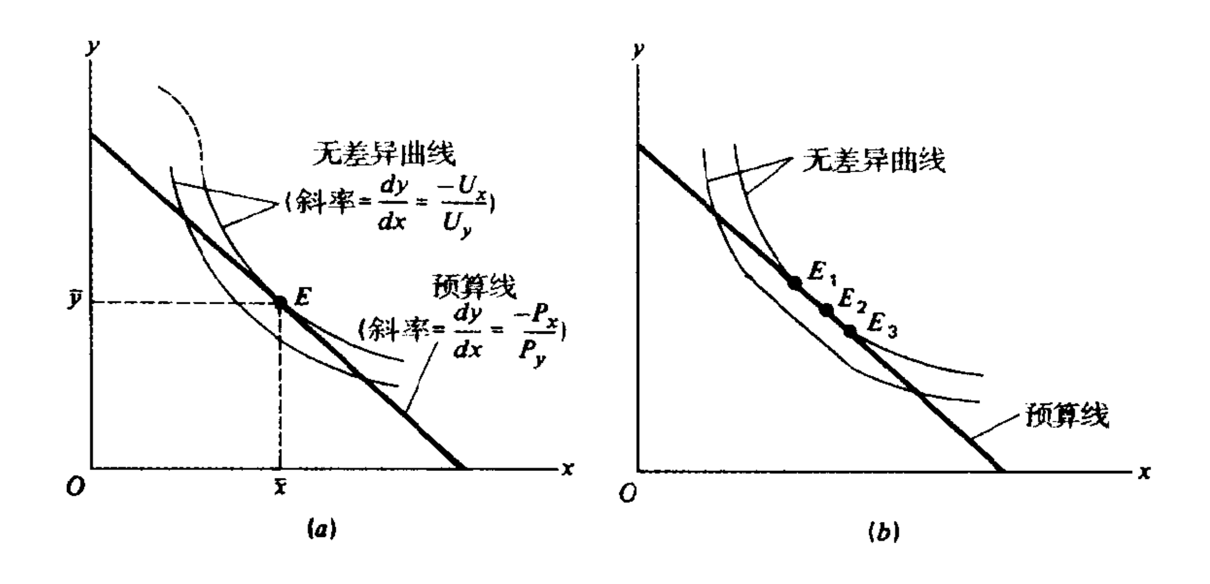



图 12.7

$$|\bar{H}| = \begin{vmatrix} 0 & P_x & P_y \\ P_x & U_{xx} & U_{xy} \\ P_y & U_{yx} & U_{yy} \end{vmatrix} = 2P_xP_yU_{xy} - P_y^2U_{xx} - P_x^2U_{yy} > 0 \tag{12.32}$$

(所有导数均在临界值 $\bar{x}$ 和 $\bar{y}$ 处计值),那么,$U$ 的稳定值必然是极大值。(12.32)中存在导数 $U_{xx}$,$U_{yy}$和 $U_{xy}$清楚表明,满足这些条件要对效用函数,因而对无差异曲线的形状施加某些限制。这些限制是什么呢?

首先考察无差异曲线的图形。我们可以证明向下倾斜的无差异曲线在切点 $E$ 严格凸。正如负的 $dy/dx(= -U_x/U_y)$可以保证无差异曲线向下倾斜一样,正的 $d^2y/dx^2$ 也可以保证无差异曲线的严格凸性。为得到表达式 $d^2y/dx^2$,我们可以将 $-U_x/U_y$ 对 $x$ 微分。但在微分时我们必须记住,$U_x$ 和 $U_y$(作为导数)都是 $x$ 和$y$ 的函数,而且在给定的无差异曲线上,$y$ 本身也是 $x$ 的函数。因而 $U_x$ 与 $U_y$ 都可以仅看作是 $x$ 的函数。因此,我们可以得到全

导数

$$\frac{d^2y}{dx^2}=\frac{d}{dx}\left(-\frac{U_x}{U_y}\right)=-\frac{1}{U_y^2}\left(U_y\frac{dU_x}{dx}-U_x\frac{dU_y}{dx}\right) \tag{12.33}$$

因 $x$ 不仅能直接影响 $U_x$ 和 $U_y$，而且通过 $y$ 作中介，也能对其产生间接影响，所以有

$$\frac{dU_x}{dx}=U_{xx}+U_{yx}\frac{dy}{dx} \qquad \frac{dU_y}{dx}=U_{xy}+U_{yy}\frac{dy}{dx} \tag{12.34}$$

其中 $dy/dx$ 是无差异曲线的斜率。现在，在切点 $E$——与讨论的二阶条件有关的唯一的点——无差异曲线斜率与预算线斜率相等；即 $dy/dx=-P_x/P_y$。因此可将(12.34)重写成 403

$$\frac{dU_x}{dx}=U_{xx}-U_{yx}\frac{P_x}{P_y} \qquad \frac{dU_y}{dx}=U_{xy}-U_{yy}\frac{P_x}{P_y} \tag{12.34'}$$

将(12.34′)代入(12.33)，并利用信息

$$U_x=\frac{U_yP_x}{P_y} \quad [由(12.31'')]$$

再提取公因子 $U_y/P_y^2$，最终可将(12.33)化为

$$\frac{d^2y}{dx^2}=\frac{2P_xP_yU_{xy}-P_y^2U_{xx}-P_x^2U_{yy}}{U_yP_y^2}=\frac{|\bar{H}|}{U_yP_y^2} \tag{12.33'}$$

显然，当满足二阶充分条件(12.32)时，(12.33′)中的二阶导数为正，相关的无差异曲线在切点为严格凸。在这里，无差异曲线在切点的严格凸性也意味着满足充分条件(12.32)。这是因为，给定处处无稳定点的无差异曲线的斜率为负，$d^2y/dx^2$ 在严格凸曲线上取零值的可能性被排除，因而现在严格凸性仅能产生正的

$d^2y/dx^2$，因此由(12.33′)，只能产生正的$|\bar{H}|$。

但是回忆一下，$|\bar{H}|$中的导数仅在临界值$\bar{x}$和$\bar{y}$处计算其值。因此，无差异曲线的严格凸性作为充分条件，仅适合于切点。如图12.7$a$中虚线所示，在离开切点$E$的地方，曲线中包含一段凹弧也是可以接受的。另一方面，如果已知效用函数为平滑、递增的拟凹函数，那么，每一无差异曲线将处处严格凸。这样的效用函数具有像图12.4$b$中那样的曲面。当这样的曲面被与$xy$平面平行的平面横切时，每次相切，我们都得到一个横截面，此截面投射到$xy$平面时，便成为一条严格拟凸，向下倾斜的无差异曲线。在此情况下，无论切点在无差异曲线的什么位置出现，二阶充分条件总是满足的。此外，无差异曲线上仅能存在一个给定预算线能够达到的唯一绝对最大效用水平的切点。这个结果与图12.6右边的菱形的结论是完全一致的。

我们曾反复提醒读者二阶充分条件并非必要条件。这里我们介绍(12.32)不成立时的效用最大化问题。如图12.7$b$所示，假设相关的无差异曲线包含一段与预算线重合的线段。那么，很明显，我们有多重极大值，因为现在一阶条件$U_x/U_y=P_x/P_y$在无差异曲线的线段部分处处得到满足，比如$E_1$、$E_2$和$E_3$等。实际上，它们都是绝对约束极大值。但因为在线段上$d^2y/dx^2$为零，由(12.33′)我们有$|\bar{H}|=0$。因此，尽管在此情况下违背二阶充分条件(12.32)，依然达到了极大化。

404

无差异曲线上出现线段表明效用曲面上存在倾斜的平面部分。当效用函数为显拟凹而不是严格拟凹时，就会出现这种情况。如图12.7$b$所示，点$E_1$、$E_2$、$E_3$均位于可实现的最高的无差异曲线上，在给定线性预算约束下产生同样的绝对最大效用。

参考图 12.6,我们注意到这一结果与左边菱形传递的信息完全一致。

## 比较静态分析

在我们的消费者模型中,价格 $P_x$ 和 $P_y$ 以及预算 $B$,均是外生的。如果假设二阶充分条件满足,那么我们可以在一阶条件(12.31)基础上分析模型的比较静态性质。将(12.31)视为一组方程 $F^j=0(j=1,2,3)$,其中函数 $F^j$ 具有连续偏导数。如(12.19)所指出的那样,这组方程内生变量雅可比行列式必与海塞加边行列式有相同的值,即 $|J|=|\bar{H}|$。因此,当满足二阶条件(12.32),$|J|$ 必为正,且在原始最优状态下不为零。因而可应用隐函数定理,且我们可将内生变量的最优值表示成外生变量的隐函数:

$$\begin{aligned}\bar{\lambda} &= \bar{\lambda}(P_x,P_y,B)\\ \bar{x} &= \bar{x}(P_x,P_y,B)\\ \bar{y} &= \bar{y}(P_x,P_y,B)\end{aligned} \tag{12.35}$$

我们知道,它们具有能够给出比较静态信息的连续偏导数。特别地,后两个函数 $\bar{x}$ 和 $\bar{y}$ 的导数描述了消费者的需求行为,它可以揭示消费者如何对价格和预算变化作出反应。但要求出那些导数,必须首先将(12.31)转换成如下均衡恒等式:

$$\begin{aligned}B-\bar{x}P_x-\bar{y}P_y &\equiv 0\\ U_x(\bar{x},\bar{y})-\bar{\lambda}P_x &\equiv 0\\ U_y(\bar{x},\bar{y})-\bar{\lambda}P_y &\equiv 0\end{aligned} \tag{12.36}$$

通过对每个恒等式依次取全微分(允许每个变量变化),并注意到 $U_{xy}=U_{yx}$,则我们得到线性方程组 405

$$-P_x d\bar{x} - P_y d\bar{y} = \bar{x}\,dP_x + \bar{y}\,dP_y - dB$$

$$-P_x d\bar{\lambda} + U_{xx} d\bar{x} + U_{xy} d\bar{y} = \bar{\lambda}\, dP_x$$

$$-P_y d\bar{\lambda} + U_{yx} d\bar{x} + U_{yy} d\bar{y} = \bar{\lambda}\, dP_y$$

(12.37)

为研究预算规模(也指消费者收入)变化的影响,令 $dP_x = dP_y = 0$,但令 $dB \neq 0$。那么,以 $dB$ 通除(12.37),并将微分的比率解释成偏导数,我们可以写出矩阵方程[①]

$$\begin{bmatrix} 0 & -P_x & -P_y \\ -P_x & U_{xx} & U_{xy} \\ -P_y & U_{yx} & U_{yy} \end{bmatrix} \begin{bmatrix} (\partial\bar{\lambda}/\partial B) \\ (\partial\bar{x}/\partial B) \\ (\partial\bar{y}/\partial B) \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \quad (12.38)$$

读者可以验证,系数矩阵中元素的排列恰与雅可比行列式 $|J|$ 中元素排列相同,$|J|$ 与 $|\bar{H}|$ 的值相同,尽管 $|\bar{H}|$ 的第一行和第一列具有 $P_x$ 和 $P_y$(而非 $-P_x$ 和 $-P_y$)。根据克莱姆法则,我们可解出三个比较静态导数,但我们将注意力集中于如下两个导数:

$$\left(\frac{\partial\bar{x}}{\partial B}\right) = \frac{1}{|J|}\begin{vmatrix} 0 & -1 & -P_y \\ -P_x & 0 & U_{xy} \\ -P_y & 0 & U_{yy} \end{vmatrix} = \frac{1}{|J|}\begin{vmatrix} -P_x & U_{xy} \\ -P_y & U_{yy} \end{vmatrix}$$

(12.39)

---

① 将(12.36)对 $B$ 求全微分,也可得到矩阵方程(12.38),但微分时要记住隐式解(12.35)。

$$\left(\frac{\partial \bar{y}}{\partial B}\right)=\frac{1}{|J|}\begin{vmatrix} 0 & -P_x & -1 \\ -P_x & U_{xx} & 0 \\ -P_y & U_{yx} & 0 \end{vmatrix}=\frac{-1}{|J|}\begin{vmatrix} -P_x & U_{xx} \\ -P_y & U_{yx} \end{vmatrix} \tag{12.40}$$

根据二阶条件，$|J|=|H|$同 $P_x$ 和 $P_y$ 一样，均为正。遗憾的是，没有 $P_x$、$P_y$ 及 $U_{ij}$ 相对大小的额外信息，我们仍难以确定这两个比较静态导数的符号。这意味着当消费者预算增加时，其最佳购买量 $\bar{x}$ 和 $\bar{y}$ 可能增加，也可能减少。当 $B$ 增加，$\bar{x}$ 减少时，产品 $x$ 被称作低质品（相对于正常品而言）。

其次，我们可以分析 $P_x$ 变化的影响。这次令 $dP_y=dB=0$，406 但使 $dP_x\neq 0$，然后以 $dP_x$ 通除(12.37)，我们得到另一个矩阵方程

$$\begin{bmatrix} 0 & -P_x & -P_y \\ -P_x & U_{xx} & U_{xy} \\ -P_y & U_{yx} & U_{yy} \end{bmatrix}\begin{bmatrix} (\partial \bar{\lambda}/\partial P_x) \\ (\partial \bar{x}/\partial P_x) \\ (\partial \bar{y}/\partial P_x) \end{bmatrix}=\begin{bmatrix} \bar{x} \\ \bar{\lambda} \\ 0 \end{bmatrix} \tag{12.41}$$

由此，得到下列比较静态导数：

$$\begin{aligned}\left(\frac{\partial \bar{x}}{\partial P_x}\right)&=\frac{1}{|J|}\begin{vmatrix} 0 & \bar{x} & -P_y \\ -P_x & \bar{\lambda} & U_{xy} \\ -P_y & 0 & U_{yy} \end{vmatrix} \\ &=\frac{-\bar{x}}{|J|}\begin{vmatrix} -P_x & U_{xy} \\ -P_y & U_{yy} \end{vmatrix}+\frac{\bar{\lambda}}{|J|}\begin{vmatrix} 0 & -P_y \\ -P_y & U_{yy} \end{vmatrix} \\ &\equiv T_1+T_2 \qquad [T_i \text{ 代表第 } i \text{ 项}]\end{aligned} \tag{12.42}$$

$$
\left(\frac{\partial \bar{y}}{\partial P_x}\right) = \frac{1}{|J|}\begin{vmatrix} 0 & -P_x & \bar{x} \\ -P_x & U_{xx} & \bar{\lambda} \\ -P_y & U_{yx} & 0 \end{vmatrix}
$$

$$
= \frac{\bar{x}}{|J|}\begin{vmatrix} -P_x & U_{xx} \\ -P_y & U_{yx} \end{vmatrix} - \frac{\bar{\lambda}}{|J|}\begin{vmatrix} 0 & -P_x \\ -P_y & U_{yx} \end{vmatrix}
$$

$$
\equiv T_3 + T_4 \qquad (12.43)
$$

如何解释这两个结果？第一个结果($\partial \bar{x} / \partial P_x$)表明 $P_x$ 的变化如何影响 $x$ 的最优购买量;因此它为我们研究消费者对 $x$ 的需求函数提供了基础。$P_x$ 变化的影响包括两项。第一项 $T_1$,运用(12.39)可重写成 $-(\partial \bar{x} / \partial B)\bar{x}$。这样,$T_1$ 似乎是 $B$(预算或收入)变化对最优购买量 $\bar{x}$ 的影响的度量,且 $\bar{x}$ 本身充当加权因子,但因此导数显然与价格变化相关,所以,必须将 $T_1$ 解释成**价格变化的收入效应**。当 $P_x$ 上升时,消费者实际收入的下降将会对 $\bar{x}$ 产生类似于 $B$ 实际下降而产生的影响,因此用 $-(\partial \bar{x} / \partial B)$表示。可以理解,商品 $x$ 在预算中的地位越突出,收入效应将越大——因此在 $T_1$ 中出现加权因子 $\bar{x}$。这个解释还可以通过以微分 $dB$ 407 $= -\bar{x}\, dP_x$ 表示的消费者有效收入的减少,来更正式地加以证明。则有

$$
\bar{x} = -\frac{dB}{dP_x}
$$

和

$$
T_1 = -\left(\frac{\partial \bar{x}}{\partial B}\right)\bar{x} = \left(\frac{\partial \bar{x}}{\partial B}\right)\frac{dB}{dP_x} \qquad (12.44)
$$

它表明,$T_1$ 是 $dP_x$ 通过 $B$ 对 $\bar{x}$ 影响的度量,亦即收入效应的度量。

如果我们现在以数量上等于 $dB$ 的现金支付来补偿消费者的

实际收入减少，那么，由于收入效应的抵销作用，比较静态导数($\partial\bar{x}/\partial P_x$)中余下的项，即 $T_2$，将度量完全由于价格变化引致的一种商品对另一种商品的替代而出现的 $\bar{x}$ 变化；即 $T_2$ 将度量 $P_x$ 变化的**替代效应**。为更清楚起见，我们回到(12.37)，考察收入补偿如何改变这种情况。当仅研究 $dP_x$ 的影响时($dP_y = dB = 0$)，(12.37)中的第一个方程可写成 $-P_x d\bar{x} - P_y d\bar{y} = \bar{x}\ dP_x$。因为消费者实际收入的减少就标示于表达式 $\bar{x}\ dP_x$ 中(顺便提一句，它仅在第一个方程中出现)，所以，对消费者进行补偿，亦即令该项等于零。若这样，(12.41)中的常数向量就应由 $\begin{bmatrix}\bar{x}\\ \bar{\lambda}\\ 0\end{bmatrix}$ 变为 $\begin{bmatrix}0\\ \bar{\lambda}\\ 0\end{bmatrix}$，且导数($\partial\bar{x}/\partial P_x$)的补偿收入形式将为

$$\left(\frac{\partial\bar{x}}{\partial P_x}\right)_{补偿} = \frac{1}{|J|}\begin{vmatrix} 0 & 0 & -P_y \\ -P_x & \bar{\lambda} & U_{xy} \\ -P_y & 0 & U_{yy} \end{vmatrix}$$

$$= \frac{\bar{\lambda}}{|J|}\begin{vmatrix} 0 & -P_y \\ -P_y & U_{yy} \end{vmatrix} = T_2$$

因此，可将(12.42)表示成如下形式

$$\left(\frac{\partial\bar{x}}{\partial P_x}\right) = T_1 + T_2 = \underbrace{-\left(\frac{\partial\bar{x}}{\partial B}\right)\bar{x}}_{收入效应} + \underbrace{\left(\frac{\partial\bar{x}}{\partial P_x}\right)_{补偿}}_{替代效应} \tag{12.42'}$$

这个结果将比较静态导数($\partial\bar{x}/\partial P_x$)分解成收入效应和替代效应两部分；此即所谓的“斯拉茨基方程”的两物品形式。

那么,如何讨论$(\partial \bar{x} / \partial P_x)$的符号呢?替代效应$T_2$明显为负,因为$|J|>0$, $\bar{\lambda}>0$[见(12.31')]。而收入效应$T_1$,按照(12.39),则是不确定的。若$T_1$为负,则$T_2$便会得到强化;在此情况下,$P_x$增加必将导致减少购买$x$,效用最大化消费者的需求曲线斜率为负。若$T_1$为正,但其值相对于$T_2$要小,则它将弱化替代效应,虽然总的结果依然以向下倾斜的需求曲线表示。但在$T_1$为正且支配$T_2$时(比如当$\bar{x}$为消费者预算中之主要项目,因而提供了一个优势加权因子时,便属于这种情况),$P_x$上升实际上会导致$x$购买的增加。具有这种需求特征的商品称作吉芬商品。当然,正常情况下,我们可以预期$(\partial \bar{x} / \partial P_x)$为负。

408

最后,我们考察(12.43)中的比较静态导数,$(\partial \bar{y} / \partial P_x) = T_3 + T_4$,它与$x$价值变化对$y$的最优购买量的交叉影响有关。$T_3$项与$T_1$项非常类似,仍可解释为收入效应。[1] 注意,这里的加权因子仍为$\bar{x}$而非$Y$,这是因为我们在研究$P_x$变化对实际收入的影响,这种影响的大小取决于$\bar{x}$(而非$\bar{y}$)在消费者预算中的相对重要性。自然,剩下的一项$T_4$,仍然是替代效应的度量。

按照(12.40),$T_3$的符号由诸如$U_{xx}$, $U_{yx}$等因子决定,如果不对模型加以限定,就不能确定其符号。但替代效应$T_4$在模型中必然为正,因为$\bar{\lambda}$, $P_x$, $P_y$及$|J|$均为正。这意味着在两商品模型中,除非为更大的负的收入效应所抵消,否则$x$价格的上升总

---

① 若读者希望更进一步确定$T_3$代表收入效应,那么,可运用(12.40)和(12.44)写出

$$T_3 = -\left(\frac{\partial \bar{y}}{\partial B}\right)\bar{x} = \left(\frac{\partial \bar{y}}{\partial B}\right)\frac{dB}{dP_x}$$

因而$T_3$表示$P_x$变化通过收入因子$B$对$\bar{y}$的影响。

会增加 $y$ 的购买。换言之，在现在的模型中（其中消费者仅有两种商品可供选择），这两种商品间存在替代关系。

尽管上面的分析仅涉及 $P_x$ 变化的影响，但我们的结论完全适用于 $P_y$ 变化的情况。我们的模型恰好是变量 $x$ 和 $y$ 完全对称的模型。所以，要推出 $P_y$ 变化的影响，只需将上面得到的结果中的 $x$ 和 $y$ 的角色互换即可。

### 价格与收入成比例的变化

考察当所有三个参数 $P_x$、$P_y$ 和 $B$ 按同一比例变化时，$\bar{x}$ 和 $\bar{y}$ 受到何种影响，也是一件有意义的工作。这种问题虽仍属比较静态学的领域，但与上面的分析有所不同，现在的研究涉及所有参数同时变化。

当两种物品价格与收入一道，均增加 $j$ 倍时，预算约束的每一项均增加 $j$ 倍，变成

$$jB - jxP_x - jyP_y = 0$$

但由于公因子 $j$ 可以消去，这个新的约束实质上与原来是一样的。而效用函数则与参数无关。因此，$x$ 和 $y$ 原来的均衡水平仍然有效，即在我们的模型中，消费者均衡状态不受价格与收入等比例变化的影响。因此，在现在的模型中，消费者没有任何“货币幻觉”。

409

这种情况可用下述方程表示

$$\bar{x}(P_x, P_y, B) = \bar{x}(jP_x, jP_y, jB)$$

$$\bar{y}(P_x, P_y, B) = \bar{y}(jP_x, jP_y, jB)$$

具有上述性质的函数 $\bar{x}$ 和 $\bar{y}$ 不是普通函数，它们是被称为齐次函数的一类特殊函数中的一例。齐次函数具有重要的经济意

义，我们下一节将对其进行考察。

## 练习 12.5

1 给定 $U=(x+2)(y+1)$ 及 $P_x=4, P_y=6, B=130$:

(*a*) 写出拉格朗日函数。

(*b*) 求最优购买水平 $\bar{x}$ 和 $\bar{y}$。

(*c*) 满足极大值二阶充分条件吗?

(*d*) (*b*)部分的答案给出比较静态信息了吗?

2 仍假设 $U=(x+2)(y+1)$，但不为价格和收入参数设定具体数值。

(*a*) 写出拉格朗日函数。

(*b*) 求 $\bar{x}$、$\bar{y}$ 及 $\bar{\lambda}$(以参数 $P_x, P_y$ 和 $B$ 表示)。

(*c*) 检验极大值的二阶充分条件。

(*d*) 令 $P_x=4, P_y=6$ 及 $B=130$，检验上面问题答案的正确性。

3 练习 12.5－2 的解($\bar{x}$ 和 $\bar{y}$)能够产生比较静态信息吗? 求出所有比较静态导数，确定其符号，并解释其经济意义。

4 由练习 12.5－2 的效用函数 $U=(x+2)(y+1)$ 和约束 $xP_x+yP_y=B$，我们已求得 $U_{ij}$ 和 $|\bar{H}|$，以及 $\bar{x}$ 和 $\bar{\lambda}$，而且回忆起来 $|J|=|\bar{H}|$。

(*a*) 将这些结果代入(12.39)和(12.40)求$(\partial\bar{x}/\partial B)$和$(\partial\bar{y}/\partial B)$。

(*b*) 将这些结果代入(12.42)和(12.43)求$(\partial\bar{x}/\partial P_x)$和$(\partial\bar{y}/\partial P_x)$。

这些结果验证了在练习 12.5－3 中所得到的结论吗?

5 评价下述论断的正确性:"如果导数$(\partial\bar{x}/\partial P_x)$为负，那么，$x$ 不可能是低质品"。

410 6 在研究 $dP_x$ 的单独影响时，(12.37)中的第一方程简化为 $-P_x d\bar{x}-P_y d\bar{y}=\bar{x}\,dP_x$，且当我们放弃 $\bar{x}\,dP_x$ 项以补偿消费者实际收入损失时，方程变成 $-P_x d\bar{x}-P_y d\bar{y}=0$，证明最后一个结果还可以通过另一种补偿方式实现，通过这种方式我们可以保持最优效用水平 $\bar{U}$(而非实际收入)不变，从而使 $T_2$ 项还可以解释成$(\partial\bar{x}/\partial P_x)_{\bar{U}=\text{常数}}$。[提示:运用(12.31″)。]

7 (*a*) 物品 $x$ 和 $y$ 的边际效用递减假设意味着无差异曲线严格凸吗?

(b) 无差异曲线的严格凸性意味着物品 $x$ 和 $y$ 的边际效用递减吗?

---

## 12.6 齐次函数

若以常数 $j$ 乘以函数的每一自变量,使函数变为原来的 $j^r$ 倍,即若

$$f(jx_1,\cdots,jx_n) = j^r f(x_1,\cdots,x_n)$$

则称此函数为 $r$ 次齐次函数。一般而言,$j$ 可取任意值。但为使上述方程有意义,$(jx_1,\cdots,jx_n)$必须在函数 $f$ 的定义域内。为此,在经济应用中常数 $j$ 通常取正值,因为绝大多数经济变量一般不取负值。

**例 1** 给定函数 $f(x,y,w)=x/y+2w/3x$,若以 $j$ 乘以每个变量,得到

$$f(jx,jy,jw) = \frac{(jx)}{(jy)} + \frac{2(jw)}{3(jx)}$$

$$= \frac{x}{y} + \frac{2w}{3x} = f(x,y,w) = j^0 f(x,y,w)$$

在这个特定的例子中,所有自变量等比例变化,函数值不受影响;或者说用倍数 $j^0(=1)$乘每个自变量,函数值不变。这样函数 $f$ 是零次齐次函数。

读者将观察到上一节末的函数 $\bar{x}$ 和 $\bar{y}$ 均是零次齐次函数。

**例 2** 若以 $j$ 乘以函数

$$g(x,y,w) = \frac{x^2}{y} + \frac{2w^2}{x}$$

的每个变量,得到

$$g(jx, jy, jw) = \frac{(jx)^2}{(jy)} + \frac{2(jw)^2}{(jx)}$$

$$= j(\frac{x^2}{y} + \frac{2w^2}{x}) = jg(x, y, w)$$

此函数为一次(或第一次)齐次函数。若以 $j$ 乘每个变量,函数值也恰好增加 $j$ 倍。

**例 3** 现在考察函数 $h(x, y, w) = 2x^2 + 3yw - w^2$。这次以 $j$ 相乘得到

$$h(jx, jy, jw) = 2(jx)^2 + 3(jy)(jw) - (jw)^2 = j^2 h(x, y, w)$$

这样函数 $h$ 是二次齐次函数;例如在本例中,所有变量加倍,函数值将增加 2 倍的平方。

## 线性齐次性

在讨论生产函数时,广泛使用的是一次齐次生产函数。它们通常被称为线性齐次函数。副词"线性地"(*linearly*)一词是修饰形容词"齐次的"。但某些作者似乎愿意使用某些易使人误解的词"线性的(*linear*)齐次函数",或者"线性和齐次的函数"等。这些词往往会传递错误的信息,使人误以为这些函数本身是线性的。由上面例 2 中的函数可以知道,一次齐次函数并不必然是线性函数。所以读者不应使用"线性的齐次函数"或"线性和齐次的函数",当然,除非所研究的函数确实是线性的。但要注意,谈及"线性的齐次性(*linear homogeneity*)"并非错误,它意指一次齐次性,因为修饰名词"齐次性"的确要求使用形容词"线性的"。

因为线性齐次函数应用的主要领域为生产理论,所以我们采用如下形式的生产函数作为我们讨论的框架:

$$Q = f(K, L) \tag{12.45}$$

无论是在宏观还是在微观水平上，线性齐次性的数学假设总是与不变规模收益的经济假设一致的，因为线性齐次性意味着所有投入（自变量）增加 $j$ 倍，总是使产出（函数值）恰好也增加 $j$ 倍。

什么性质表示线性齐次生产函数的特征呢？

**性质 I** 给定线性齐次生产函数 $Q=f(K,L)$，劳动的平均物质产品（$APP_L$）和资本的平均物质产品（$APP_K$）可以表示成资本劳动比率 $k\equiv K/L$ 的函数。412

为证明这一点，我们将(12.45)中的每个自变量都乘以一个因子 $j=1/L$。根据线性齐次性，这将使产出由 $Q$ 变成 $jQ=Q/L$。相应地，(12.45)的右侧将变成

$$f(\frac{K}{L},\frac{L}{L}) = f(\frac{K}{L},1) = f(k,1)$$

因为原函数中的变量 $K$ 和 $L$ 已分别被 $k$ 和 1 所替换（无论 $K$ 和 $L$ 何时出现），实际上方程右边变成仅是资本劳动比率的函数，比如 $\phi(k)$，它是一个单一变量 $k$ 的函数，尽管两个自变量 $K$ 和 $L$ 实际上包含在变量 $k$ 中。令左右两边相等，有

$$APP_L \equiv \frac{Q}{L} = \phi(k) \tag{12.46}$$

于是求得 $APP_K$ 的表达式为

$$APP_K \equiv \frac{Q}{K} = \frac{Q}{L}\frac{L}{K} = \frac{\phi(k)}{k} \tag{12.47}$$

因为两个平均产量仅取决于 $k$，线性齐次性意味着，只要比率 $K/L$ 保持不变（无论 $K$ 和 $L$ 的绝对水平是多少），那么，平均产量也将保持不变。因此，尽管生产函数是一次齐次性的，但 $APP_L$ 和 $APP_K$ 均是变量 $K$ 和 $L$ 的零次齐次函数，因为 $K$ 和 $L$ 的等比例变化（保持 $k$ 为常数）将不会改变平均产量的大小。

**性质** II　给定线性齐次生产函数 $Q=f(K,L)$，边际物质产品 $MPP_L$ 和 $MPP_K$ 可以仅表示成 $k$ 的函数。

为求得边际产量，我们将总产量写成

$$Q = L\phi(k) \quad [由(12.46)] \qquad (12.45')$$

然后将 $Q$ 对 $K$ 和 $L$ 求微分。为此，我们先求出如下两个初步结果以供使用

$$\frac{\partial k}{\partial K} = \frac{\partial}{\partial K}\left(\frac{K}{L}\right) = \frac{1}{L} \qquad \frac{\partial k}{\partial L} = \frac{\partial}{\partial L}\left(\frac{K}{L}\right) = \frac{-K}{L^2} \qquad (12.48)$$

413 微分的结果为

$$\begin{aligned} MPP_K \equiv \frac{\partial Q}{\partial K} &= \frac{\partial}{\partial K}[L\phi(k)] \\ &= L\frac{\partial \phi(k)}{\partial K} = L\frac{d\phi(k)}{dk}\frac{\partial k}{\partial K} \qquad [链式法则] \\ &= L\phi'(k)\left(\frac{1}{L}\right) = \phi'(k) \quad [由(12.48)] \quad (12.49) \end{aligned}$$

$$\begin{aligned} MPP_L \equiv \frac{\partial Q}{\partial L} &= \frac{\partial}{\partial L}[L\phi(k)] \\ &= \phi(k) + L\frac{\partial \phi(k)}{\partial L} \qquad [积的微分法则] \\ &= \phi(k) + L\phi'(k)\frac{\partial k}{\partial L} \qquad [链式法则] \\ &= \phi(k) + L\phi'(k)\frac{-K}{L^2} \qquad [由(12.48)] \\ &= \phi(k) - k\phi'(k) \qquad (12.50) \end{aligned}$$

它确实表明 $MPP_L$ 和 $MPP_K$ 仅是 $k$ 的函数。

像平均产量一样，只要资本劳动比率不变，边际产量亦将保持不变，它们是变量 $K$ 和 $L$ 的零次齐次函数。

**性质 III （欧拉定理）** 若 $Q=f(K,L)$为线性齐次函数，则

$$K\frac{\partial Q}{\partial K}+L\frac{\partial Q}{\partial L}\equiv Q$$

证明

$$K\frac{\partial Q}{\partial K}+L\frac{\partial Q}{\partial L}=K\phi'(k)+L[\phi(k)-k\phi'(k)]$$

[由(12.49)),(12.50)]

$$=K\phi'(k)+L\phi(k)-K\phi'(k) \qquad [k\equiv K/L]$$

$$=L\phi(k)=Q \qquad [由(12.45')]$$

注意，对任意 $K$ 和 $L$ 值，此结论均成立。这是为什么上式被写成恒等式的原因。这个性质表明线性齐次函数的值总可以表示成若干项的和，其中每一项是一个自变量与 $Q$ 对该自变量一阶偏导数之乘积，而不管实际采用的两种投入是多少。但应仔细区分恒等式 $K\frac{\partial Q}{\partial K}+L\frac{\partial Q}{\partial L}\equiv Q$[欧拉定理仅应用于$Q=f(K,L)$为不变规模收益的情况]和方程 $dQ=\frac{\partial Q}{\partial K}dK+\frac{\partial Q}{\partial L}dL$[任意函数 $Q$ 的全微分]的区别。 414

在经济上，这个性质表示在不变规模收益条件下，如果每种投入要素按其边际产品数量获得报酬，那么，每种投入要素获得的分配份额之和恰好等于总产出，或者说，净经济利润为零。因为这种情况是对完全竞争条件下长期均衡的描述，所以曾有人认为，在经济学中只有线性齐次函数才有意义。当然，情况并非如此。在长期均衡中的零经济利润是由于新进入的厂商和已存在厂商之间的竞争所致，而与实际存在的生产函数的具体性质无关。因此，具有产品为任何和所有生产要素$(K,L)$所分配完毕的生产函数，并非是必须的。而且当要素市场存在不完全竞争时，每种要素所获得

的报酬可能不等于边际产品，结果，欧拉定理可能与分配情况无关。但由于我们知道线性齐次生产函数具有许多有特色的数学性质，因而便于应用。

### 柯布—道格拉斯生产函数

在经济分析中广泛使用的一种具体生产函数是柯布—道格拉斯生产函数：

$$Q = AK^{\alpha}L^{1-\alpha} \tag{12.51}$$

其中 $A$ 为正常数，$\alpha$ 为正分数。我们这里首先要考察的是此函数的一般化形式

$$Q = AK^{\alpha}L^{\beta} \tag{12.52}$$

其中 $\beta$ 为另一个正分数，它可能等于，也可能不等于 $1-\alpha$。此函数的一些主要特性是：(1) 它是 $(\alpha+\beta)$ 次齐次函数；(2) 在 $(\alpha+\beta)$ 等于 1 的特定情况下，它是线性齐次函数；(3) 对于正的 $K$ 和 $L$ 值，其等产量曲线是一条斜率处处为负的严格凸曲线；(4) 对于正的 $K$ 和 $L$，它是严格拟凹的。

它的齐次方程可以很容易从这一事实中看到：如果将 $K$ 和 $L$ 分别变成 $jK$ 和 $jL$，产出就会变成

$$A(jK)^{\alpha}(jL)^{\beta} = j^{\alpha+\beta}(AK^{\alpha}L^{\beta}) = j^{\alpha+\beta}Q$$

由此很容易看出此函数的齐次性，即此函数是 $(\alpha+\beta)$ 次齐次函数。在 $(\alpha+\beta)=1$ 的情况下，将有不变规模收益，因为函数是线性齐次的。（但要注意，此函数不是线性的，要避免将其与“线性的齐次的函数”及“线性的和齐次的函数”相混淆。）其等产量曲线斜率为负且为严格凸可以由导数 $dK/dL$ 和 $d^2K/dL^2$ 的符号（或者 $dL/dK$ 及 $d^2L/dK^2$ 的符号）来验证。对于正的产出 $Q_0$，

415

(12.52)可以写成

$$AK^{\alpha}L^{\beta} = Q_0 \quad (A, K, L, Q_0 > 0)$$

对两边取自然对数并移项,求得

$$\ln A + \alpha \ln K + \beta \ln L - \ln Q_0 = 0$$

此方程将 $K$ 定义为 $L$ 的隐函数[①]。根据隐函数法则和对数的求导法则,我们有:

$$\frac{dK}{dL} = -\frac{\partial F/\partial L}{\partial F/\partial K} = -\frac{(\beta/L)}{(\alpha/K)} = -\frac{\beta K}{\alpha L} < 0$$

则由此得

$$\frac{d^2K}{dL^2} = \frac{d}{dL}\left(-\frac{\beta K}{\alpha L}\right) = -\frac{\beta}{\alpha}\frac{d}{dL}\left(\frac{K}{L}\right) = -\frac{\beta}{\alpha}\frac{1}{L^2}\left(L\frac{dK}{dL} - K\right) > 0$$

这些导数的符号确定了对于正的 $K$ 和 $L$ 值,在 $LK$ 平面上(任意)等产量曲线斜率处处为负且严格凸。当然,这是我们对一个对于正的 $K$ 和 $L$ 为严格拟凹的函数的唯一期望。对于此函数的严格拟凹特征,可参见 12.4 节例 5,那里讨论了一个类似的函数。

现在考察 $\alpha + \beta = 1$(柯布—道格拉斯生产函数本身)的情况,以验证前面提出的线性齐次性的三个性质。首先,此特定情况下中的总产量可以表示成

$$Q = AK^{\alpha}L^{1-\alpha} = A\left(\frac{K}{L}\right)^{\alpha}L = LAk^{\alpha} \qquad (12.51')$$

其中表达式 $Ak^{\alpha}$ 是前面使用的一般表达式 $\phi(k)$ 的具体形式。因此,平均产量为

$$\mathrm{APP_L} = \frac{Q}{L} = Ak^{\alpha}$$

---

① 因为 $F$(左边表达式)具有连续偏导数,而且对正的 $K$ 值,$\partial F/\partial K = \alpha/K \neq 0$,所以满足隐函数定理的条件。

$$\text{APP}_K = \frac{Q}{K} = \frac{Q}{L}\frac{L}{K} = \frac{Ak^\alpha}{k} = Ak^{\alpha-1} \qquad (12.53)$$

二者现在均仅为 $k$ 的函数。

416　其次,微分 $Q = AK^\alpha L^{1-\alpha}$,得到边际产量:

$$\frac{\partial Q}{\partial K} = A\alpha K^{\alpha-1}L^{-(\alpha-1)} = A\alpha\left(\frac{K}{L}\right)^{\alpha-1} = A\alpha k^{\alpha-1}$$

$$\frac{\partial Q}{\partial L} = AK^\alpha(1-\alpha)L^{-\alpha} = A(1-\alpha)\left(\frac{K}{L}\right)^\alpha = A(1-\alpha)k^\alpha \qquad (12.54)$$

它们也都仅是 $k$ 的函数。

最后,我们运用(12.54)验证欧拉定理:

$$\begin{aligned} K\frac{\partial Q}{\partial K} + L\frac{\partial Q}{\partial L} &= KA\alpha k^{\alpha-1} + LA(1-\alpha)k^\alpha \\ &= LAk^\alpha\left[\frac{K\alpha}{Lk} + 1 - \alpha\right] \\ &= LAk^\alpha[\alpha + 1 - \alpha] = LAk^\alpha = Q \quad [\text{由}(12.51')] \end{aligned}$$

在线性齐次道格拉斯生产函数中,指数 $\alpha$ 和 $(1-\alpha)$ 具有重要的经济意义。如果假设每种投入按其边际产量获得报酬,那么,资本在总产量中的相对份额将为

$$\frac{K(\partial Q/\partial K)}{Q} = \frac{KA\alpha k^{\alpha-1}}{LAk^\alpha} = \alpha$$

类似地,劳动的相对份额为

$$\frac{L(\partial Q/\partial L)}{Q} = \frac{LA(1-\alpha)k^\alpha}{\text{LAk}^\alpha} = 1 - \alpha$$

因此,每种投入变量的指数表示该投入在总产量中的相对份额。从另一个角度看,我们也可以将每种投入变量的指数解释成产出对该投入的偏弹性。这是因为上面给出的资本份额表达式等价于

$\frac{\partial Q/\partial K}{Q/K} \equiv \varepsilon_{QK}$，类似地，上面的劳动份额表达式等价于 $\varepsilon_{QL}$。

那么，常数 $A$ 是何含义呢？对于给定的 $K$ 和 $L$ 值，$A$ 将对产出水平 $Q$ 产生与 $A$ 大小成比例的影响。因此可以视 $A$ 为一个**效率参数**，即作为反映技术水平的指标。

## 结果的推广

我们已就生产函数的特定情况讨论了线性齐次性，但是只要将变量 $K$、$L$ 及 $Q$ 重新作适当解释，那么，上面所提到的性质在其它情况下依然成立。

还可以将上述结果推广到两个以上变量的情况。有线性齐次 417 函数

$$y = f(x_1, x_2, \cdots, x_n)$$

将每个变量以 $x_1$ 除(即乘以 $1/x_1$)，得到结果

$$y = x_1 \phi\left(\frac{x_2}{x_1}, \frac{x_3}{x_1}, \cdots, \frac{x_n}{x_1}\right) \quad [\text{一次齐次性}]$$

它与(12.45′)具有可比性。而且欧拉定理也可以很容易推广成如下形式

$$\sum_{i=1}^{n} x_i f_i \equiv y \qquad [\text{欧拉定理}]$$

其中原函数 $f$ 的偏导数(即 $f_i$)如同两变量情况一样，依然是变量 $x_i$ 的零次齐次函数。

实际上，上述结果还可以轻松推广至 $r$ 次齐次函数的情况。首先，根据齐次性的定义，现在可以写成

$$y = x_1^r \phi\left(\frac{x_2}{x_1}, \frac{x_3}{x_1}, \cdots, \frac{x_n}{x_1}\right) \qquad [r\ \text{次齐次性}]$$

修正的欧拉定理以下述形式出现

$$\sum_{i=1}^{n} x_i f_i \equiv ry \quad [\text{欧拉定理}]$$

其中乘积常数 $r$ 附属于右边自变量 $y$。最后,原函数 $f$ 的偏导数 $f_i$,是变量 $x_i$ 的 $(r-1)$ 次齐次函数。因此,读者可以看到,线性齐次性只不过是 $r=1$ 时的一种特例。

## 练习 12.6

---

1 确定下列函数是否是齐次函数。若是,是几次齐次函数?

(a) $f(x,y)=\sqrt{xy}$　　(d) $f(x,y)=2x+y+3\sqrt{xy}$

(b) $f(x,y)=(x^2-y^2)^{1/2}$　　(e) $f(x,y,w)=\dfrac{xy^2}{w}+2xw$

(c) $f(x,y)=x^3-xy+y^3$　　(f) $f(x,y,w)=x^4-5yw^3$

2 证明(12.45)不仅可以表示成 $Q=L\phi\left(\dfrac{K}{L}\right)$ 还可以表示成

$Q=K\phi\left(\dfrac{L}{K}\right)$。

418

3 在固定规模收益情况下,由欧拉定理推出:

(a) 当 $MPP_K=0$ 时,$APP_L=MPP_L$。

(b) 当 $MPP_L=0$ 时,$APP_K=MPP_K$。

4 在(12.46)至(12.50)的基础上,检验在固定规模收益条件下,下列结论是否成立。

(a) 可以对作为自变量的 $k(=K/L)$ 绘出一条 $APP_L$ 曲线($k$ 作为横轴)。

(b) $APP_L$ 曲线的斜率可以度量 $MPP_K$。

(c) $APP_L$ 曲线的向量半径的斜率可以度量 $APP_K$。

(d) $MPP_L=APP_L-k(MPP_K)=APP_L-k(APP_L$ 的斜率)。

5 运用(12.53)和(12.54)证明科布—道格拉斯生产函数服从上一题 $b$、$c$、$d$ 部分所描述的关系。

6　给定生产函数 $Q = AK^{\alpha}L^{\beta}$，证明

(*a*) $(\alpha + \beta) > 1$ 意味着规模收益递增。

(*b*) $(\alpha + \beta) < 1$ 意味着规模收益递减。

(*c*) $\alpha$ 和 $\beta$ 分别是产出对资本与劳动投入的偏弹性。

7　令产出是三种投入的函数：即 $Q = AK^{a}L^{b}N^{c}$。

(*a*) 此函数是齐次的吗？

(*b*) 在何条件下会存在不变规模收益？递增的规模收益？

(*c*) 如果按边际产品的数量获得支付，求投入 $N$ 应获得的产品份额。

8　令 $Q = g(K, L)$ 是二次齐次生产函数。

(*a*) 写出表示此函数二次齐次函数性质的方程。

(*b*) 根据(12.45′)的风格，以 $\phi(k)$ 写出 $Q$ 的表达式。

(*c*) 求 $MPP_K$ 函数。$MPP_K$ 仍像线性齐次性的情况一样，只是 $k$ 的函数吗？

(*d*) $MPP_K$ 是 $K$ 和 $L$ 的齐次函数吗？若是，是几次的？

---

## 12.7　投入的最小成本组合

作为约束最优化的另一个例子，我们讨论求产出为给定水平 $Q_0$ 的最小成本投入组合问题。这个给定的产出水平也许是为某顾客的订货而生产的。这里我们使用一般生产函数，但后面，也将涉及齐次生产函数。

### 一阶条件

假设有一平滑的具有两种可变投入的生产函数 $Q = Q(a, b)$，其中 $Q_a$ 和 $Q_b > 0$，并假设两种投入价格均为外生的（但仍舍 419 弃零下标），这样我们就可以构建一个问题：使成本

$$C = aP_a + bP_b$$

最小化，满足产出约束

$$Q(a,b) = Q_0$$

因此，拉格朗日函数为

$$Z = aP_a + bP_b + \mu[Q_0 - Q(a,b)]$$

为满足最小化 $C$ 的一阶条件，投入水平(选择变量)必定满足如下联立方程

$$Z_\mu = Q_0 - Q(a,b) = 0$$

$$Z_a = P_a - \mu Q_a = 0$$

$$Z_b = P_b - \mu Q_b = 0$$

上面第一个方程只不过是约束条件的重述，后两个方程意味着条件

$$\frac{P_a}{Q_a} = \frac{P_b}{Q_b} = \mu \tag{12.55}$$

在最优投入组合点，投入价格与边际产出的比率对每一投入必定相同。因为此比率度量该投入的单位边际产出的支出数量，所以拉格朗日乘数可以被解释成为最优状态下的产品边际成本。当然，这个解释同前面在(12.16)中发现的拉格朗日乘数的最优值是比较静态影响的度量完全一致。此影响即约束方程中的常数对目标函数最优值的影响，即 $\bar{\mu} = (\S \bar{C} / \S Q_0)$，其中符号 § 表示它是一个偏全导数。

方程(12.55)还可以写成这种形式

$$\frac{P_a}{P_b} = \frac{Q_a}{Q_b} \tag{12.55'}$$

读者可将其与(12.31″)相比较。以这种形式出现，一阶条件还可以按照等产量曲线和等成本曲线来解释。正如我们在(11.36)中已经知道的那样，$Q_a/Q_b$ 是等产量曲线斜率的负值；即，它是$a$对

$b$ 边际技术替代率的度量($MRTS_{ab}$)。在现在的模型中,产出水平设定为 $Q_0$,因此仅涉及一条等产量曲线,如图 12.8 所示,其斜率为负。

而比率 $P_a/P_b$ 则表示等成本曲线(可将其与消费者理论中的预算线相比较)斜率的负值。等成本曲线定义为具有相同总成本的投入组合的轨迹。它可以下述方程表示

$$C_0 = aP_a + bP_b \quad 或 \quad b = \frac{C_0}{P_b} - \frac{P_a}{P_b}a$$

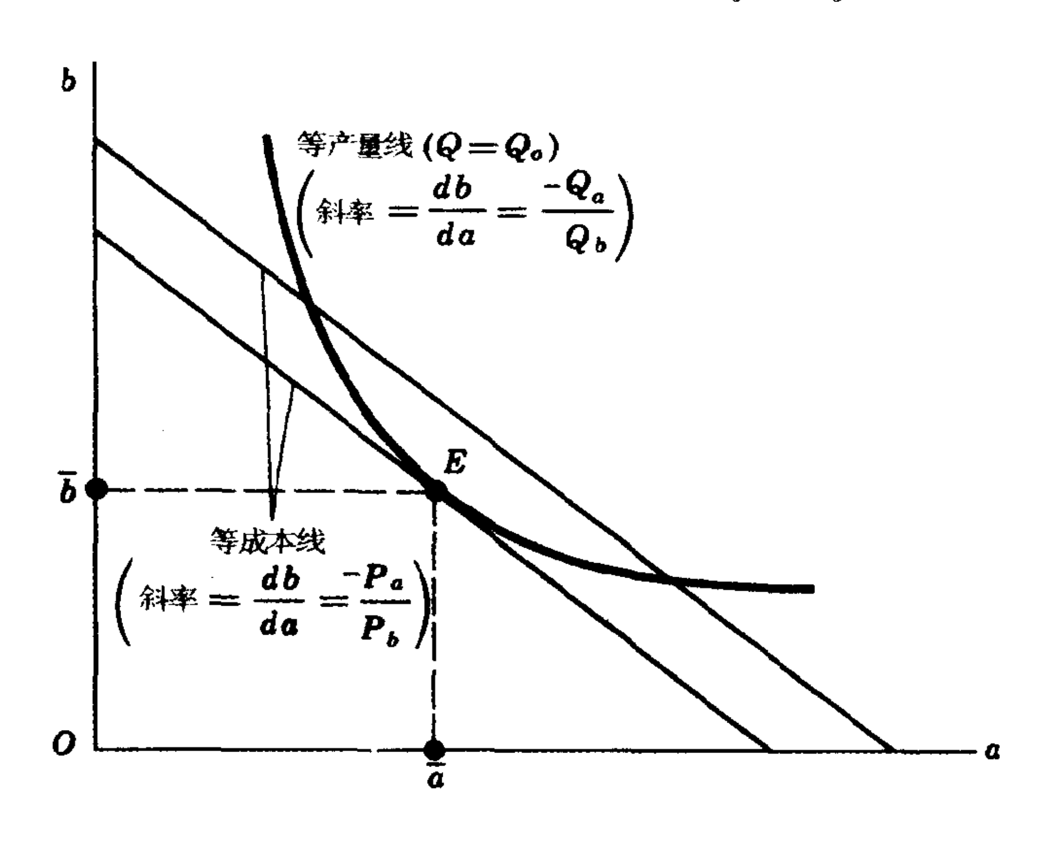



图 12.8

其中 $C_0$ 代表(参数)成本数字。因此,如图 12.8 所示,当在 $ab$ 平面绘出时,产出斜率为 $-P_a/P_b$(和纵截距为 $C_0/P_b$)的一族直线。420 所以,两个比率相等相当于等产量曲线的斜率与所选定的等成本曲线的斜率相等。因为我们必须维持在给定等产量曲线上,这个条件使我们得到切点 $E$ 和投入组合($\bar{a}$,$\bar{b}$)。

## 二阶条件

在一阶条件满足后，只要有负的海塞加边行列式，即有

$$|\bar{H}| = \begin{vmatrix} 0 & Q_a & Q_b \\ Q_a & -\mu Q_{aa} & -\mu Q_{ab} \\ Q_b & -\mu Q_{ba} & -\mu Q_{bb} \end{vmatrix}$$
$$= \mu(Q_{aa}Q_b^2 - 2Q_{ab}Q_aQ_b + Q_{bb}Q_a^2) < 0$$

那么，便足以保证最小成本。因为 $\mu$(边际成本)的最优值为正，这样使条件符号为在 $E$ 点计值时，括号中的表达式为负。

由(11.40)，我们可以回忆起等产量曲线的曲率是以二阶导数来表示的

$$\frac{d^2b}{da^2} = \frac{-1}{Q_b^3}(Q_{aa}Q_b^2 - 2Q_{ab}Q_aQ_b + Q_{bb}Q_a^2)$$

其中出现了与前面相同的括号内的表达式。由于 $Q_b$ 为正，二阶充分条件的满足意味着 $d^2b/da^2$ 为正——即等产量曲线为严格凸——在切点处。在这里，等产量曲线的严格凸性也意味着满足421二阶充分条件。因为等产量曲线斜率为负，严格凸性只能意味着正的 $d^2b/da^2$(在等产量曲线的稳定点处，$d^2b/da^2$ 为零也是可能的)，进而又可以保证 $|\bar{H}|<0$。但还应当记住，充分条件在切点处 $|\bar{H}|<0$(因而等产量曲线严格凸性)实质上并非 $C$ 最小化的必要条件。具体地，即使当等产量曲线是非严格凸时，在类似于图12.7$b$ 的多重最小值情况下，在每个最小值有 $d^2b/da^2=0$ 和 $|\bar{H}|=0$，$C$ 仍可能取最小值。

在讨论效用最大化模型(见12.5节)时，我们曾指出，平滑、递增、严格拟凹的效用函数可以在 $xy$ 平面产生处处严格凸的向下倾

斜的无差异曲线。因为等产量曲线的概念几乎与无差异曲线是一致的,[①] 我们可以类推,平滑、递增、严格拟凹的生产函数 $Q=Q(a,b)$ 也可以在 $ab$ 平面产生处处严格凸的向下倾斜的等产量曲线。如果假设存在这样的生产函数,那么,显然总是满足二阶充分条件。而且,很明显,由此而得到的 $C$ 将是唯一的绝对约束极小值。

## 扩展线

现在转向此模型的比较静态分析方面。假定投入价格**固定**,我们让 $Q_0$ 持续递增(升至越来越高的等产量曲线),追寻这对最小成本组合 $\bar{b}/\bar{a}$ 的影响。当然等产量曲线的每次移动将与更高的等成本曲线产生一个新的切点。这些切点的轨迹被称作厂商的**扩展线**,它描述生产不同产出水平 $Q_0$ 所需的最小成本组合。两种可能的扩展线形状如图 12.9 所示。

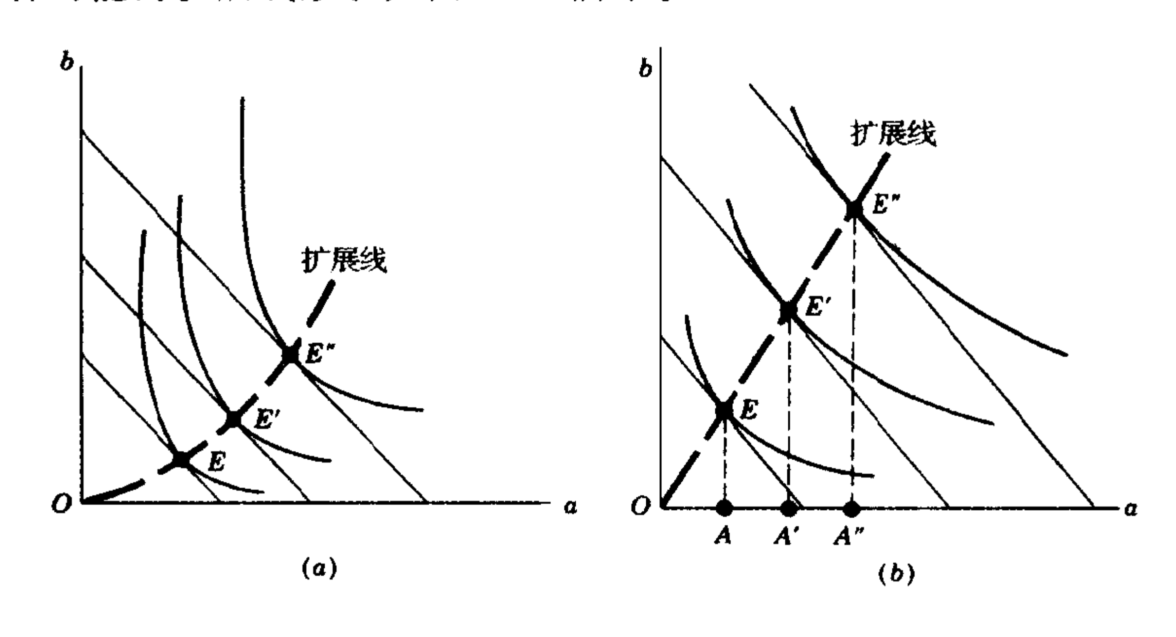



图 12.9

① 二者实质上均是“等值”曲线。它们的不同仅在于所应用的领域不同。无差异曲线应用于消费模型,而等产量曲线应用于生产模型。

若我们假设等产量曲线具有严格凸性，因而满足二阶条件，则扩展线可直接从条件(12.55′)推导出来。我们用一般形式的科布—道格拉斯生产函数对此进行描述。

条件(12.55′)要求投入价格比率与边际产量比率相等。对于函数 $Q = Aa^{\alpha}b^{\beta}$，这意味着扩展线的每一点必定满足

$$\frac{P_a}{P_b} = \frac{Q_a}{Q_b} = \frac{A\alpha a^{\alpha-1}b^{\beta}}{Aa^{\alpha}\beta b^{\beta-1}} = \frac{\alpha b}{\beta a} \tag{12.56}$$

422 这意味着最优投入比率应为

$$\frac{b}{a} = \frac{\beta P_a}{\alpha P_b} = a \text{ 常数} \tag{12.57}$$

因为 $\alpha$、$\beta$、投入价格均为常数。因此，扩展线上所有点必定表示相同的固定的投入比率；即扩展线必定为一条自原点出发的直线。这种关系描述在图 12.9$b$ 中，其中在不同的切点($AE/OA$，$AE'/OA'$，$AE''/OA''$)的投入比率都相等。

线性扩展线是一般化的科布—道格拉斯生产函数的特征，而与 $\alpha+\beta$ 是否等于 1 无关，因为结果(12.57)的推导并不依赖于 $\alpha+\beta=1$的假设。事实上，任何齐次生产函数(不必一定是科布—道格拉斯生产函数)对于每一组投入价格都会产生线性扩展线，原因如下：如果此生产函数是 $r$ 次齐次的，那么其边际产量函数 $Q_a$ 和 $Q_b$ 必是投入 $a$ 和 $b$ 的$(r-1)$次齐次函数，因此两种投入增加 $j$ 倍将使 $Q_a$ 和 $Q_b$ 的值变化 $j^{r-1}$倍，但比率 $Q_a/Q_b$ 没有任何变化。因此，若一个特定投入组合$(a_0, b_0)$在给定投入价格下满足一阶条件 $P_a/P_b = Q_a/Q_b$，那么，投入组合$(ja_0, jb_0)$必定也满足——恰如图 12.9$b$ 线性扩展线所描述的那样。

尽管**任意**齐次生产函数均可产生线性扩展线，但齐次性的次

数不同，确使在对扩展线进行解释时产生很大的不同。在图12.9$b$中，我们绘出的距离$OE$等于$EE'$，从而$E'$的刻度为$E$的两倍。现在生产函数若是一次齐次的，那么，$E'$的产出必定为$E$的二倍$(2^1=2)$。但若生产函数是二次齐次的，$E'$的产出则是$E$的四倍$(2^2=4)$。这样，对于$Q=1, Q=2,\cdots\cdots$等等产量曲线的间隔，对于不同的齐次性而言是有很大不同的。 423

## 位似函数

我们已经解释了给定一组投入价格，(任意次)齐次生产函数会产生线性扩展线。但线性扩展线并不仅为齐次生产函数所特有；一种更一般类型的函数，即所谓位似函数，也能产生线性扩展线。

位似函数是一个形式如下的复合函数：

$$H = h[Q(a,b)] \quad [h'(Q) \neq 0] \tag{12.58}$$

其中$Q(a,b)$是$r$次齐次的。尽管$H=H(a,b)$源于齐次函数，但一般而言它并不是变量$a$和$b$的齐次函数。然而，$H(a,b)$的扩展线，像$Q(a,b)$的扩展线一样，都是线性的。这一结论的关键之点在于，在$ab$平面上的任意给定点，$H$的等量线与$Q$的等产量线具有相同的斜率：

$$\begin{aligned} H\text{ 等量线的斜率} &= -\frac{H_a}{H_b} = -\frac{h'(Q)Q_a}{h'(Q)Q_b} \\ &= -\frac{Q_a}{Q_b} = Q\text{ 等量线的斜率} \end{aligned} \tag{12.59}$$

现在$Q(a,b)$扩展线的线性意味着条件

$$\text{对于任意给定的}\frac{b}{a}, -\frac{Q_a}{Q_b} = \text{常数},$$

且此条件也意味着$Q(a,b)$的扩展线为线性。但根据(12.59)，立即有

对于任意给定的$\frac{b}{a}$，$-\frac{H_a}{H_b}$ = 常数 (12.60)

这确定了 $H(a,b)$ 也会产生线性扩展线。

位似性是一个比齐次性更一般的概念。事实上，每个齐次函数均是位似函数族中的一员，但位似函数则可能不是齐次函数中的一员。其实齐次函数总是位似函数可由(12.58)看出。在(12.58)中，如果我们令函数 $H=h(Q)$ 取特定形式 $H=Q$（有 $h'(Q)=dH/dQ=1$），则函数 $Q$ 与函数 $H$ 相同，它显然是位似函数。关于位似函数不一定是齐次函数在下面例 2 中阐述。

在定义位似函数 $H$ 时，我们在(12.58)中设定 $h'(Q)\neq 0$。这可以使我们免于在(12.59)中以零作分母。设定 $h'(Q)\neq 0$ 仅是数学处理上的要求，而从经济角度考虑，则要求更强的限制 $h'(Q)>0$。因为如果 $H(a,b)$ 像 $Q(a,b)$ 一样作为生产函数，即如果 $H$ 表示产出，那么，$H_a$ 和 $H_b$ 应分别像 $Q(a,b)$ 函数中的 $Q_a$ 和 $Q_b$ 那样按同样方向变化。因此，$H(a,b)$ 需被限定为 $Q(a,b)$ 的单调递增变换。

424

位似函数，包括其特例齐次函数，具有这样一种重要性质：最优投入水平对产出水平的偏弹性对所有投入而言都是相同的。为证明这一点，回顾一下，位似函数扩展线的线性意味最优投入比率 $\bar{b}/\bar{a}$ 不受外生产出水平 $H_0$ 的影响。因此 $\partial(\bar{b}/\bar{a})/\partial H_0=0$，或者

$$\frac{1}{\bar{a}^2}\left(\bar{a}\frac{\partial \bar{b}}{\partial H_0}-\bar{b}\frac{\partial \bar{a}}{\partial H_0}\right)=0 \quad [\text{商的微分法则}]$$

以 $\bar{a}^2 H_0$ 通乘并重排，得到

$$\frac{\partial \bar{a}}{\partial H_0}\frac{H_0}{\bar{a}}=\frac{\partial \bar{b}}{\partial H_0}\frac{H_0}{\bar{b}} \quad \text{或} \quad \varepsilon_{\bar{a}H_0}=\varepsilon_{\bar{b}H_0}$$

这正是我们上面所下的结论。

**例 1** 令 $H=Q^2$,其中 $Q=Aa^{\alpha}b^{\beta}$。因为 $Q(a,b)$是齐次的,且 $h'(Q)=2Q$ 对于正的产出为正,所以对于 $Q>0$,$H(a,b)$是位似函数。我们将证明它满足(12.60)。首先通过替代,我们有:

$$H = Q^2 = (Aa^{\alpha}b^{\beta})^2 = A^2a^{2\alpha}b^{2\beta}$$

因此 $H$ 的等产量线的斜率可以表示成

$$-\frac{H_a}{H_b} = -\frac{A^2 2\alpha a^{2\alpha-1}b^{2\beta}}{A^2a^{2\alpha}2\beta b^{2\beta-1}} = -\frac{\alpha b}{\beta a} \qquad (12.61)$$

此结果满足(12.60)且意味着线性扩展线。将(12.61)与(12.56)比较也可以证明函数 $H$ 满足(12.59)。

在本例中,$Q(a,b)$是$(\alpha+\beta)$次齐次函数。而且 $H(a,b)$也是齐次函数,但次数为 $2(\alpha+\beta)$。但从规律上看,位似函数不一定是齐次函数。

**例 2** 令 $H=e^Q$,其中 $Q=Aa^{\alpha}b^{\beta}$。因为 $Q(a,b)$是齐次函数,且 $h'(Q)=e^Q$ 为正,所以 $H(a,b)$是位似函数。由此函数

$$H(a,b) = \exp(Aa^{\alpha}b^{\beta})$$

很易求得

$$-\frac{H_a}{H_b} = -\frac{A\alpha a^{\alpha-1}b^{\beta}\exp(Aa^{\alpha}b^{\beta})}{Aa^{\alpha}\beta b^{\beta-1}\exp(Aa^{\alpha}b^{\beta})} = -\frac{\alpha b}{\beta a}$$

当然,这个结果与例 1 中的(12.61)是一致的。

但这次位似函数**不是**齐次函数,因为

425

$$H(ja,jb) = \exp[A(ja)^{\alpha}(jb)^{\beta}] = \exp(Aa^{\alpha}b^{\beta}j^{\alpha+\beta})$$
$$= [\exp(Aa^{\alpha}b^{\beta})]^{j^{\alpha+\beta}} = [H(a,b)]^{j^{\alpha+\beta}} \neq j^rH(a,b)$$

## 替代弹性

比较静态分析的另一方面是分析比率 $P_a/P_b$ 变化对能够生

产出相同的给定产出水平 $Q_0$(即维持在相同的等产量曲线上)的最小成本投入组合 $\bar{b}/\bar{a}$ 的影响。

当(外生的)投入价格比率 $P_a/P_b$ 提高时,一般可以预期最优投入比率 $\bar{b}/\bar{a}$ 也会提高,因为现在相对便宜的投入 $b$ 将倾向于替代投入 $a$。替代的方向是清楚的,但替代的程度如何呢?投入的替代程度可以用以 $\sigma$(小写希腊字母 sigma,表示"替代")表示的所谓替代弹性来度量:

$$\sigma \equiv \frac{(\bar{b}/\bar{a})\text{的相对变化}}{(P_a/P_b)\text{的相对变化}}$$

$$= \frac{\dfrac{d(\bar{b}/\bar{a})}{\bar{b}/\bar{a}}}{\dfrac{\mathrm{d}(P_a/P_b)}{P_a/P_b}} = \frac{\dfrac{d(\bar{b}/\bar{a})}{d(P_a/P_b)}}{\dfrac{\bar{b}/\bar{a}}{\mathrm{P_a}/P_b}} \tag{12.62}$$

$\sigma$ 的值位于 0 至 $\infty$ 之间;$\sigma$ 越大,两种投入的替代能力越强。$\sigma=0$ 这种极限情况表示两种投入必须按固定比例互补使用。$\sigma$ 为无限大的另一种极限情况表明两种投入具有完全的替代性。注意,若 $(\bar{b}/\bar{a})$ 被视为 $(P_a/P_b)$ 的函数,那么,弹性 $\sigma$ 仍将是边际函数对平均函数的比率。

---

① 还有一种表示 $\sigma$ 的方式。因为在切点,我们总有

$$\frac{P_a}{P_b} = \frac{Q_a}{Q_b} = \mathrm{MRTS}_{ab}$$

等价地,替代弹性可以定义为

$$\sigma = \frac{(\bar{b}/\bar{a})\text{的相对变化}}{\mathrm{MRTS}_{ab}\text{的相对变化}} = \frac{\dfrac{d(\bar{b}/\bar{a})}{\bar{b}/\bar{a}}}{\dfrac{\mathrm{d}(Q_a/Q_b)}{Q_a/Q_b}} = \frac{\dfrac{d(\bar{b}/\bar{a})}{d(Q_a/Q_b)}}{\dfrac{\bar{b}/\bar{a}}{\mathrm{Q_a}/Q_b}} \tag{12.62'}$$

下面我们说明如何计算一般化的科布—道格拉斯生产函数的替代弹性。我们前面知道，对于这种情况，最小成本投入组合由 426

$$\left(\frac{\bar{b}}{\bar{a}}\right) = \frac{\beta}{\alpha}\left(\frac{P_a}{P_b}\right) \quad [\text{由}(12.57)]$$

所规定。此方程的形式为 $y = ax$，对于这个方程 $dy/dx$（边际）和 $y/x$（平均）均等于常数 $a$，即

$$\frac{d(\bar{b}/\bar{a})}{d(P_a/P_b)} = \frac{\beta}{\alpha} \quad \text{和} \quad \frac{\bar{b}/\bar{a}\ F}{P_a/P_b} = \frac{\beta}{\alpha}$$

将这些值代入(12.62)，立即求得 $\sigma = 1$；即一般化的科布—道格拉斯生产函数是以不变的单位替代弹性为特征的。注意，此结论的推导完全不依赖于 $\alpha + \beta = 1$ 的假设。因此，生产函数 $Q = Aa^{\alpha}b^{\beta}$ 的替代弹性即使 $\alpha + \beta \neq 1$，也是 1。

## CES 生产函数

最近，人们常采用另一种形式的生产函数，虽然这种生产函数仍以不变替代弹性(constant elasticity of substitution，简写为 CES)为特征，但 $\sigma$ 值也可能不等于 1。[①] 此函数的方程，即所谓 CES 生产函数，是

$$Q = A[\delta K^{-\rho} + (1-\delta)L^{-\rho}]^{-1/\rho}$$

$$(A > 0; 0 < \delta < 1; -1 < \rho \neq 0) \qquad (12.63)$$

其中 $K$ 和 $L$ 表示两种生产要素，$A$，$\delta$ 和 $\rho$（小写希腊字母 rho）为三个参数。参数 $A$（效率参数）与柯布—道格拉斯生产函数中的系数 $A$ 起相等作用，是一个用于反映技术水平的指标。参数 $\delta$

---

① K.J.阿罗，H.B.钱纳里，B.S.明哈斯，R.M.索罗："资本劳动替代和经济效率"，载《经济学与统计学评论》，1961 年秋季号，225—250 页。

(分配参数)类似于科布—道格拉斯生产函数中的 $\alpha$,与产品中的相对要素份额有关。参数 $\rho$(替代参数,在柯布—道格拉斯生产函数中并无其对应物)是不变替代弹性的值,后面将证明这一点。

首先,我们考察这种函数的一次齐次函数。若以 $jK$ 和 $jL$ 分别替代 $K$ 和 $L$,产出将由 $Q$ 变成

$$
\begin{aligned}
& A[\delta(jK)^{-\rho}+(1-\delta)(jL)^{-\rho}]^{-1/\rho} \\
& = A\{j^{-\rho}[\delta K^{-\rho}+(1-\delta)L^{-\rho}]\}^{-1/\rho} \\
& = (j^{-\rho})^{-1/\rho}Q = jQ
\end{aligned}
$$

结果,CES 生产函数,像线性齐次生产函数一样,表现为不变规模收益,符合应用欧拉定理的条件,且具有平均产量和边际产量函数,它们都是 $K$ 和 $L$ 的零次齐次函数。

427 我们也应注意到,CES 生产函数产生的等产量曲线斜率总是为负,且对正的 $K$ 和 $L$ 值为严格凸。为证明这一点,我们首先求边际产量 $Q_L$ 和 $Q_K$ 的表达式。以符号[…]表示 $[\delta K^{-\rho}+(1-\delta L^{-\rho}]$ 的简写形式,有

$$
\begin{aligned}
Q_L \equiv \frac{\partial Q}{\partial L} & = A\left(-\frac{1}{\rho}\right)[\cdots]^{-(1/\rho)-1}(1-\delta)(-\rho)L^{-\rho-1} \\
& = (1-\delta)A[\cdots]^{-(1+\rho)/\rho}L^{-(1+\rho)} \\
& = (1-\delta)\frac{A^{1+\rho}}{A^{\rho}}[\cdots]^{-(1+\rho)/\rho}L^{-(1+\rho)} \\
& = \frac{(1-\delta)}{A^{\rho}}\left(\frac{Q}{L}\right)^{1+\rho} > 0 \quad [\text{由}(12.63)]
\end{aligned}
\tag{12.64}
$$

类似地,

$$
Q_K \equiv \frac{\partial Q}{\partial K} = \frac{\delta}{A^{\rho}}\left(\frac{Q}{K}\right)^{1+\rho} > 0 \tag{12.65}
$$

它们是对正的 $K$ 和 $L$ 而定义的。因此等产量曲线($K$ 为纵轴,$L$ 为横轴)为

$$\frac{dK}{dL} = -\frac{Q_L}{Q_K} = -\frac{(1-\delta)}{\delta}\left(\frac{K}{L}\right)^{1+\rho} < 0 \quad [\text{见}(11.36)] \tag{12.66}$$

则很容易验证 $d^2K/dL^2>0$(其证明留与读者作练习)意味着等产量曲线对正的 $K$ 和 $L$ 为严格凸的。

还可以证明,CES 生产函数对正的 $K$ 和 $L$ 是拟凹的。对(12.64)及(12.65)进一步微分,可以证明此函数的二阶导数符号如下

$$Q_{LL} = \frac{\partial}{\partial L}Q_L = \frac{(1-\delta)(1+\rho)}{A^\rho}\left(\frac{Q}{L}\right)^\rho \frac{Q_L L - Q}{L^2} < 0$$

[根据欧拉定理,$Q_L L - Q<0$]

$$Q_{KK} = \frac{\partial}{\partial K}Q_K = \frac{\delta(1+\rho)}{A^\rho}\left(\frac{Q}{K}\right)^\rho \frac{Q_K K - Q}{K^2} < 0$$

[根据欧拉定理,$Q_K K - Q<0$]

$$Q_{KL} = Q_{LK} = \frac{(1-\delta)(1+\rho)}{A^\rho}\left(\frac{Q}{L}\right)^\rho \frac{Q_K}{L} > 0$$

对于正的 $K$ 和 $L$ 成立的这些导数符号,使我们得以检验拟凸性的充分条件(12.26)。正如读者可以证明的那样

$$|B_1| = -Q_K^2 < 0$$

和 $$|B_2| = 2Q_K Q_L Q_{KL} - Q_K^2 Q_{LL} - Q_L^2 Q_{KK} > 0$$

因此,CES 生产函数对于正的 $K$ 和 $L$ 是拟凹的。 428

最后,我们运用边际产量(12.64)和(12.65)求 CES 生产函数的替代弹性。为满足最小成本组合条件 $Q_L/Q_K = P_L/P_K$,其中 $P_L$ 和 $P_K$ 分别表示劳动的价格(工资率)和资本的价格(资本品的

租金)，必须有

$$\frac{1-\delta}{\delta}\left(\frac{K}{L}\right)^{1+\rho}=\frac{P_L}{P_K}\quad[见(12.66)]$$

因此，最优投入比率为(引入简写符号 $c$)

$$\left(\frac{\bar{K}}{\bar{L}}\right)=\left(\frac{\delta}{1-\delta}\right)^{1/(1+\rho)}\left(\frac{P_L}{P_K}\right)^{1/(1+\rho)}\equiv c\left(\frac{P_L}{P_K}\right)^{1/(1+\rho)}$$

(12.67)

取$(\bar{K}/\bar{L})$为$(P_L/P_K)$的函数，我们求得有关的边际函数和平均函数如下：

$$边际函数=\frac{d(\bar{K}/\bar{L})}{d(P_L/P_K)}=\frac{c}{1+\rho}\left(\frac{P_L}{P_K}\right)^{1/(1+\rho)-1}$$

$$平均函数=\frac{\bar{K}/\bar{L}}{P_L/P_K}=c\left(\frac{P_L}{P_K}\right)^{1/(1+\rho)-1}$$

因此，替代弹性为：①

$$\sigma=\frac{边际函数}{平均函数}=\frac{1}{1+\rho}\qquad(12.68)$$

此式表明，$\sigma$ 是一个常数，其大小取决于参数 $\rho$ 的值：

$$\left.\begin{array}{r}-1<\rho<0\\ \rho=0\\ 0<\rho<\infty\end{array}\right\}\Rightarrow\left\{\begin{array}{l}\sigma>1\\ \sigma=1\\ \sigma<1\end{array}\right.$$

---

① 当然，首先对(12.67)两边取对数，我们也可以得到这一结

$$\ln\left(\frac{\bar{K}}{\bar{L}}\right)=\ln c+\frac{1}{1+\rho}\ln\left(\frac{P_L}{P_K}\right)$$

然后应用(10.28)中的弹性公式，得到：

$$\sigma=\frac{d(\ln\bar{K}/\bar{L})}{d(\ln P_L/P_K)}=\frac{1}{1+\rho}$$

## 作为 CES 生产函数特例的柯布—道格拉斯生产函数

在上面最后一个结果中，那个中间情况 $\rho=0$ 导致单位替代弹性；而这正是科布—道格拉斯生产函数的特征。这表明，(线性齐次)柯布—道格拉斯生产函数是(线性齐次)CES 生产函数的一个特征。困难的是，CES 生产函数〔如(12.63)给出的那样〕，当 $\rho=0$时是没有意义的，因为以零作除数是不可能的。然而，可以证明，当 $\rho\to0$ 时，CES 生产函数趋近于科布—道格拉斯生产函数。 429

为证明这一点，我们需借助于所谓的**洛比达法则**的方法。此法则与函数 $f(x)=\frac{m(x)}{n(x)}$当 $x\to a$ 时的极限的计算有关(其中 $a$ 可以为有限，也可以为无限)，当分子 $m(x)$和分母 $n(x)$有如下情况之一时：(1) 当 $x\to a$ 时，二者均趋于零，因而产生一个 0/0 形式的表达式；或者(2) 当 $x\to a$ 时，二者均趋于 $\pm\infty$，因而产生一个形式为$\infty/\infty$(或$\infty/-\infty$，或$-\infty/\infty$，或$-\infty/-\infty$)的表达式。尽管 $f(x)$的极限在上述两种情况下难以直接计算，但运用下述公式却可计算出其极值

$$\lim_{x\to a}\frac{m(x)}{n(x)}=\lim_{x\to a}\frac{m'(x)}{n'(x)} \quad [\text{洛比达法则}] \qquad (12.69)$$

**例 3** 求当 $x\to1$ 时，$(1-x^2)(1-x)$的极限。这里，当 $x\to1$ 时，$m(x)$和 $n(x)$均趋近于零，因此，符合上面的情况(1)。因为 $m'(x)=-2x$，$n'(x)=-1$，我们可以写出

$$\lim_{x\to1}\frac{1-x^2}{1-x}=\lim_{x\to1}\frac{-2x}{-1}=\lim_{x\to1}2x=2$$

此答案与 6.4 节用另一种方法得到的结果是一致的。

**例 4** 求当 $x\to\infty$ 时，$(2x+5)/(x+1)$ 的极限。当 $x\to\infty$ 时，本例中的 $m(x)$ 和 $n(x)$ 均为无穷大，因而是上面情况(2)的一个例子。因为 $m'(x)=2$，$n'(x)=1$，所以可以写成

$$\lim_{x\to\infty}\frac{2x+5}{x+1}=\lim_{x\to\infty}\frac{2}{1}=2$$

同样，这个答案与 6.4 节例 3 以另一种方法得到的答案是一致的。

可以证明，(12.69) 右边的表达式仍可能像左边的表达式一样，出现 0/0 或 $\infty/\infty$ 的形式。在这种情况下，我们可以再应用洛比达法则，即考察当 $x\to a$ 时，$m''(x)/n''(x)$ 的极限，并取此极限为我们的答案。也可能出现这种情况：虽然我们希望求其极值的给定函数 $f(x)$ 最初不是 $m(x)/n(x)$ 的形式，即在取极值时不是 0/0 或 $\infty/\infty$ 的形式，但经适当变换仍可使 $f(x)$ 适于应用法则 (12.69)。这种可能性可以通过求 CES 生产函数 (12.63) 当 $\rho\to0$ 时的极值来加以说明。现在将 (12.63) 视作函数 $Q(\rho)$。

正如已给定的那样，$Q_\rho$ 并不是 $m(\rho)/n(\rho)$ 的形式。但以 $A$ 除 (12.63) 的两边，并取自然对数，我们确实可以得到符合洛比达法则形式的函数，即

430

$$\ln\frac{Q}{A}=\frac{-\ln[\delta K^{-\rho}+(1-\delta)L^{-\rho}]}{\rho}\equiv\frac{m(\rho)}{n(\rho)}\qquad(12.70)$$

进而，当 $\rho\to0$ 时，可求得 $m(\rho)\to\ln[\delta+1-\delta]=-\ln1=0$，及 $n(\rho)\to0$。因此，洛比达法则可用于求 $\ln(Q/A)$ 的极限。若如此，$Q$ 的极限也可求出：因为 $Q/A=e^{\ln(Q/A)}$，所以 $Q=Ae^{\ln(Q/A)}$，由此得

$$\lim Q=\lim Ae^{\ln(Q/A)}=Ae^{\lim\ln(Q/A)}\qquad(12.71)$$

由 (12.70)，我们首先按洛比达法则的需要，求出 $m'(\rho)$ 和 $n'(\rho)$。后者仅为 $n'(\rho)=1$。前者是

$$m'(\rho)=\frac{-1}{[\delta K^{-\rho}+(1-\delta)L^{-\rho}]}\frac{d}{d\rho}[\delta K^{-\rho}+(1-\delta)L^{-\rho}] \quad [链式法则]$$

$$=\frac{-[-\delta K^{-\rho}\ln K-(1-\delta)L^{-\rho}\ln L]}{[\delta K^{-\rho}+(1-\delta)L^{-\rho}]} \quad [由(10.21')]$$

因此，由洛比达法则，有

$$\lim_{\rho\to 0}\ln\frac{Q}{A}=\lim_{\rho\to 0}\frac{m'(\rho)}{n'(\rho)}$$

$$=\frac{\delta\ln K+(1-\delta)\ln L}{1}=\ln(K^{\delta}L^{1-\delta})$$

根据这一结果，当 $e$ 被自乘到 $\lim_{\rho\to 0}\ln(Q/A)$ 次幂时，结果仅是 $K^{\delta}L^{1-\delta}$。因此，由(12.71)，最终得到这一结果

$$\lim_{\rho\to 0}Q=AK^{\delta}L^{1-\delta}$$

它表明，当 $\rho\to 0$ 时，CES 生产函数确实趋向于科布—道格拉斯生产函数。

## 练习 12.7

---

1 假设图 12.9$b$ 中的等产量曲线是由一个特定的齐次生产函数 $Q=Q(a,b)$推导出来的。注意，$OE=EE'=E'E''$，若函数 $Q$ 是

($a$) 一次齐次的 ($b$) 二次齐次的

那么，由三个等产量曲线所表示的产出水平间的比率是多少？

2 对于一般化的科布—道格拉斯生产函数，如果我们对比率 $P_a/P_b$ 绘出比率 $\bar{b}/\bar{a}$，会得到何种类型的曲线？这个结果依赖于假设 $\alpha+\beta=1$ 吗？从图上读出此曲线的替代弹性。

3 对于所有正的投入水平，CES 生产函数对每一投入具有收益递减的特征吗？

4 证明在 CES 生产函数等产量曲线上，$d^2K/dL^2>0$。

431 5 ($a$) 对于 CES 生产函数,若每一生产要素按其边际产量获得报酬,那么,劳动获得的产品份额与资本获得的产品份额的比率是多少?$\delta$ 越大意味着资本获得的相对份额越大吗?

($b$) 对于科布—道格拉斯生产函数,劳动获得的产品份额与资本获得的产品份额的比率依赖于比率 $K/L$ 吗?

6 ($a$)CES 生产函数排除了 $\rho=-1$ 的情况。但若 $\rho=-1$,对于正的 $K$ 和 $L$,等产量曲线一般会是什么形状?

($b$) 对于 $\rho=-1$,$\sigma$ 有意义吗?当 $\rho\rightarrow-1$ 时,$\sigma$ 的极限是多少?

($c$) 解释上述结果的经济含义。

7 证明:把 CES 生产函数写成 $Q=A[\delta K^{-\rho}+(1-\delta)L^{-\rho}]^{-r/\rho}$,其中 $r>0$是一个新参数,我们可以引入递增的规模收益和递减的规模收益。

8 求下列极限

($a$) $\lim\limits_{x\to4}\dfrac{x^2-x-12}{x-4}$  ($c$) $\lim\limits_{x\to0}\dfrac{5^x-e^x}{x}$

($b$) $\lim\limits_{x\to0}\dfrac{e^x-1}{x}$  ($d$) $\lim\limits_{x\to\infty}\dfrac{\ln x}{x}$

9 运用洛比达法则,证明

($a$) $\lim\limits_{x\to\infty}\dfrac{x^n}{e^x}=0$  ($b$) $\lim\limits_{x\to0^+}x\ln x=0$  ($c$) $\lim\limits_{x\to0^+}x^x=1$

---

## 12.8 结束语

在本篇,我们研究了最优化的基本方法。我们已走过了多少有些艰难的行程:(1) 从单一选择变量的情况到更一般的 $n$ 个选择变量的情况;(2) 从多项式目标函数到指数和对数目标函数;(3) 从非约束极值到约束极值等形式。

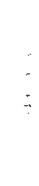
我们的讨论大多数由“古典”最优化方法所构成,它主要依靠的是微积分,各阶导数是其基本工具。以微积分方法求最优化的

一个缺陷是其内在的短见性质。虽然导数或微分形式的一阶和二阶条件一般不难确定相对或局部极值，但通常仍需收寻其它信息以识别绝对或整体极值。我们对凹性、凸性、拟凹性和拟凸性的讨论旨在为从相对极值领域迈向绝对极值领域提供一些有益的垫脚石。

微积分法的一个更严重的局限性在于它不能应付约束条件为不等式形式的情况。由于这个原因，如在效用最大化模型的预算约束是以这种形式表述的：总支出恰好等于（而不是“小于或等于”）特定的收入和。换言之，微积分法的局限性使其排斥了消费者储蓄其部分收入的选择。而且正是由于同样的原因，我们没有像一般经济常识所表明的那样，明确地将选择变量限定为非负。432 事实上，我们只能像模型设定的那样使用不等式（如 $Q_a > 0$ 和 $Q_{aa} < 0$ 等等）。这些不等式在确定解的数学符号时起作用，但它们本身不是数学运算的对象。

当研究代表“非古典”方法的数学规划（线性规划和非线性规划）时，我们将处理不等式约束问题。这个问题将留待本书第六篇讨论。同时，为使读者全方位理解经济分析的各个领域——静态学→比较静态学→动态学——我们将在第五篇介绍动态分析方法。这也是数学方法所特别推崇的，因为动态分析所需的数学方法与我们刚刚学过的微积分方法有密切的联系。

但对于那些急于转向数学规划的读者而言，完全可以跨过第五篇，直接进入第六篇，这不会遇到方法上的障碍。



# 第五篇

# 动 态 分 析

435

# 第13章　动态经济学与积分学

把动态学这一术语应用于经济分析时，在不同的时间以及对不同的经济学家，都有不同的含义。[①] 然而在今天的标准用法中，它是指这样一种分析类型：其目的是探寻和研究变量的具体时间路径，或者是确定在给定的充分长的时间内，这些变量是否会趋向收敛于某一(均衡)值。这方面的研究是非常重要的，因为它可以弥补静态学和比较静态学的严重不足。在比较静态学中，我们总是武断地假设：经济调节过程不可避免地导致均衡。而在动态分析中，我们直接面对均衡的"可实现性"问题，而不是假设它必然能够实现。

动态分析的一个显著特征是确定变量的时间，这就把时间因素明确纳入分析范围。有两种方式可以做到这一点：我们可以将时间视为连续变量，也可以将其视为离散变量。在前一种情况下，变量在每一时点都要发生某些变化(如在连续计算复利时那样)；436 而在后一种情况下，变量仅在某一时段内才发生某些变化(如仅在每六个月末才计入利息)。这两个不同的时间概念在不同的内容中各具优势。

我们将首先讨论连续时间的情况，它与积分学及微分方程等

① 弗里兹·马契拉普："静态学与动态学：众说纷纭的术语"，载《南方经济杂志》，1959年10月，pp 91—110；重印于其《经济语义学论文集》，Prentice－Hall出版公司，1963年版，pp 9—42。

数学方法有关。在后面16、17章,我们再讨论离散时间的情况,那里将运用差分方程的方法。

## 13.1 动态学与积分

一般而言,静态模型中的问题是要求出满足某些特定均衡条件的内生变量的值。把静态学应用于最优化模型时,任务变成求使目标函数最大化(或最小化)的选择变量的值——而一阶条件充当均衡条件。与此相对照的是,动态模型涉及的问题是,在已知变化模式的基础上(比如,给定瞬时变化率),描述某些变量的变化时间路径。

举个例子或许会使问题更清楚。假定已知人口规模 $H$ 随时间以速率

$$\frac{dH}{dt} = t^{-1/2} \tag{13.1}$$

变化。则我们要求的是:人口 $H = H(t)$ 的何种时间路径可以产生(13.1)的变化率?

读者将会认识到,如果我们起初便知道函数 $H = H(t)$,那么便可以通过微分求得 $dH/dt$。但我们现在面临的问题恰恰相反:要从已知的导数求出原函数,而不是从原函数求出其导数。在数学上,我们现在需要与微分法或微分学完全相反的方法。

这种方法称作积分法或积分学,我们下面将对其进行研究。现在,我们满足于如下观察:$H(t) = 2t^{1/2}$确实有形式为(13.1)的导数,因此显然可以作为我们的问题的解。但麻烦的是,还存在类似的函数,如 $H(t) = 2t^{1/2} + 15$ 或 $H(t) = 2t^{1/2} + 99$,更一般地,

$$H(t) = 2t^{1/2} + c \ (c = \text{任意常数}) \qquad (13.2)$$

它们均与(13.1)有完全相同的导数。这样就不能确定唯一的时间路径,除非常数值 $c$ 能以某种方式确定下来。为此,模型必须以所谓初始条件或边界条件的形式,引入额外的信息。

437 如果我们知道初始人口 $H(0)$(即 $H$ 在 $t=0$)时的值,假设 $H(0)=100$),则常数 $c$ 的值可以确定了。令(13.2)中的 $t=0$,得到

$$H(0) = 2(0)^{1/2} + c = c$$

但若 $H(0)=100$,则 $c=100$,且(13.2)变成

$$H(t) = 2t^{1/2} + 100 \qquad (13.2')$$

其中的常数不再是任意的。更一般地,对于任意给定初始人口 $H(0)$,时间路径将为

$$H(t) = 2t^{1/2} + H(0) \qquad (13.2'')$$

因此,在现在的例子中,任意时点的人口规模由初始人口 $H(0)$ 与另一个包含时间变量大的项的和组成。这个时间路径的确描述了变量 $H$ 随时间变化的过程,因此确实构成了此动态模型的解。[方程(13.1)也是大的函数。为什么它不能被视为模型的解呢?]

人口问题的例子虽然简单,但却揭示了动态经济学问题的实质:给定变量随时间变化的行为模式,设法求出描述变量时间路径的函数。在此过程中,我们将遇到一个或多个任意常数,但我们若有充分的形式为初始条件的额外信息,就有可能确定那些任意常数的值。

相对简单的问题,比如上面给出的例子,解可用积分方法求出。积分是一种由已知导函数反求原函数的方法。在更复杂的情况中,我们还可以借助于被称作微分方程的方法;微分方程是一个与积分密切相关的数学分支。因为微分方程被定义为包含微分或

导数表达式的任意方程，所以(13.1)必定是一个微分方程。因此，通过求其解，我们实际上已经解出了一个微分方程，尽管它是一个极其简单的微分方程。

现在我们开始积分学的基本概念。因为我们在讨论微积分时使用 $x$(而非 $t$)作自变量，为对称起见，这里仍将使用 $x$。但为方便计，我们在这里的讨论中分别以 $F(x)$和 $f(x)$表示原函数和导函数，而不是使用一撇“′”来区分它们。

## 13.2 不定积分

### 积分的性质

我们曾提过，积分是微分的逆过程。若给定原函数 $F(x)$，对其微分得到导数 $f(x)$，假设可以得到适当的信息以确定在积分过程中产生的任意常数，那么，我们就可以“积分”$f(x)$以求得 $F$ 438 $(x)$。函数 $F(x)$被称作 $f(x)$的**积分**或反导数。因此这两种运算过程类似于研究家谱的两种方法：**积分**就是追溯函数 $f(x)$的家系或出身，而微分则是寻找 $F(x)$的后裔。但要注意其区别：尽管可微原函数总是产生一个后代，即唯一的导数 $f(x)$，但导数 $f(x)$的积分，则可能追溯到无数个可能父母；因为若 $F(x)$是 $f(x)$的积分，则正如我们在(13.2)所看到的那样，$F(x)$加上任意常数也是 $f(x)$的积分。

我们需要一个特殊符号来表示所求的 $f(x)$对 $x$ 的积分。标准符号为

$$\int f(x)dx$$

左边加长的 $S$ 形符号(它有和的含义,后面再解释)称作积分符,而 $f(x)$ 部分则称作被积函数,$dx$ 部分,类似于微分算子 $d/dx$ 中的 $dx$ 部分,则提示我们运算是对 $x$ 进行的。但读者也可以把 $f(x)dx$ 视为一个整体,并将其解释成原函数 $F(x)$ 的微分[即 $dF(x)=f(x)dx$]。那么,前面的积分符号可以看作是逆微分过程的指示符。有了这个新的符号便可以写出

$$\frac{d}{dx}F(x) = f(x) \Rightarrow \int f(x)dx = F(x) + c \qquad (13.3)$$

其中任意积分常数 $c$,用以表示被积函数的多重亲系。

更具体地,积分 $\int f(x)$ 被称作 $f(x)$ 的不定积分(与下一节要讨论的定积分相对照),因为它没有确定的数值。由于它等于 $F(x)+c$,其值随 $x$ 的变化而变化(即便 $c$ 是确定的)。因此同导数一样,不定积分本身是 $x$ 的函数。

### 不定积分的基本法则

同存在一些求导法则一样,我们也可以提出一些积分法则。正如所预期的那样,积分法则极大地依赖于我们已熟悉的微分法则。例如,由如下幂函数的导数公式

$$\frac{d}{dx}\left(\frac{x^{n+1}}{n+1}\right) = x^n \qquad (n \neq -1)$$

我们看到表达式 $x^{n+1}/(n+1)$ 是导函数 $x^n$ 的原函数。因此,用439 它们代替(13.3)中的 $F(x)$ 和 $f(x)$,我们可以表述如下积分法则。

**法则 I (幂函数积分法则)**

$$\int x^n dx = \frac{1}{n+1}x^{n+1} + c \ (n \neq -1)$$

**例 1**　求 $\int x^3dx$ 。这里,我们有 $n=3$。因此有

$$\int x^3dx = \frac{1}{4}x^4 + c$$

**例 2**　求 $\int xdx$。因为 $n=1$,所以有

$$\int xdx = \frac{1}{2}x^2 + c$$

**例 3**　求 $\int 1dx$。为求此积分,我们回忆一下,$x^0=1$;因此,在幂函数积分法则中可令 $n=0$,得到

$$\int 1dx = x + c$$

[有时 $\int 1dx$ 被写成 $\int dx$,因为 $1dx=dx$。]

**例 4**　求 $\int \sqrt{x^3}dx$。因 $\sqrt{x^3}=x^{3/2}$,所以有 $n=3/2$,因此,

$$\int \sqrt{x^3}dx = \frac{x^{5/2}}{\frac{5}{2}} + c = \frac{2}{5}\sqrt{x^5} + c$$

**例 5**　求 $\int \frac{1}{x^4}dx$,$(x \neq 0)$。因 $1/x^4=x^{-4}$,所以有 $n=-4$。这样,积分为

$$\int \frac{1}{x^4}dx = \frac{x^{-4+1}}{-4+1} + c = -\frac{1}{3x^3} + c$$

注意,积分结果的正确性总可以通过微分来检验;若积分是正确的,那么,积分的导数必然等于被积函数。

我们已给出简单指数和对数函数的求导公式为

$$\frac{d}{dx}e^x = e^x \quad 和 \frac{d}{dx}\ln x = \frac{1}{x} \quad (x > 0)$$

由此可得出两个基本的积分法则。

**法则 II （指数函数的积分法则）**

$$\int e^x dx = e^x + c$$

**法则 III （对数函数的积分法则）**

$$\int \frac{1}{x} dx = \ln x + c \quad (x > 0)$$

令人感兴趣的是，法则 III 中所包含的被积函数是 $1/x = x^{-1}$，它是幂函数 $x^n$ 中当 $n = -1$ 的特殊形式。这个特殊的被积函数难以按幂函数积分法则求积分，但现在却完全可以通过对数函数的积分法则来处理。

440

如上所述，对数函数积分法则被置以约束 $x>0$，因为 $x$ 的非正值不存在对数。该法则的可以处理 $x$ 负值的更一般的公式为

$$\int \frac{1}{x} dx = \ln|x| + c \quad (x \neq 0)$$

此式如 $(d/dx)\ln x = 1/x$ 一样，也意味着$(d/dx)\ln|x| = 1/x$。读者应当确信，以$|x|$(具有限制 $x \neq 0$)代替 $x$(具有限制 $x>0$)丝毫不影响公式的正确性。

还应指出，作为一个符号，积分$\int \frac{1}{x} dx$ 还可以写成$\int \frac{dx}{x}$。作为法则 II 和法则 III 的变形，我们还有如下两个法则。

**法则 IIa**

$$\int f'(x) e^{f(x)} dx = e^{f(x)} + c$$

**法则 IIIa**

$$\int \frac{f'(x)}{f(x)} dx = \ln f(x) + c \qquad [f(x) > 0]$$

$$\text{或} \qquad \ln|f(x)| + c \qquad [f(x) \neq 0]$$

在求导法则(10.20)中可以找到这两个法则的基础。

## 运算法则

上面给出的三个法则充分说明了积分的所有法则所依赖的精神实质,即每个法则总是对应于一个求导公式,而且每个积分公式后面都附加一个任意常数(尽管通过运用给定的边界条件,可以使其确定下来),以表示原函数的整个亲系都可以产生给定的被积函数。

但为处理更复杂的被积函数,我们还需求出如下两个对积分有帮助的运算法则:

**法则 IV(和的积分)**:有限个函数的和的积分,等于这些函数 441 积分之和。对于两个函数的情况,这意味着

$$\int [f(x)+g(x)]dx = \int f(x)dx + \int g(x)dx$$

此法则是下述事实之必然结果

$$\underbrace{\frac{d}{dx}[F(x)+G(x)]}_{A} = \underbrace{\frac{d}{dx}F(x)+\frac{d}{dx}G(x)}_{B} = \underbrace{f(x)+g(x)}_{C}$$

由于 $A=C$,基于(13.3),可以写出

$$\int [f(x)+g(x)]dx = F(x)+G(x)+c \qquad (13.4)$$

但由 $B=C$,可得

$$\int f(x)dx = F(x)+c_1 \text{ 和} \int g(x)dx = G(x)+c_2$$

这样,通过相加可得到

$$\int f(x)dx + \int g(x)dx = F(x)+G(x)+c_1+c_2 \qquad (13.5)$$

因为常数 $c, c_1, c_2$ 均为任意值,我们可令 $c = c_1 + c_2$,则方程(13.4)与(13.5)的右边相等,这样,其左边必然也相等。这证明了法则 IV。

**例 6** 求 $\int (x^3 + x + 1)dx$。由法则 IV,此积分可以表示成三个积分的和:$\int x^3 dx + \int x dx + \int 1 dx$。因为此三个积分的值已分别在例 1、2、3 中求出,只需将其结果合起来,便得到:

$$\int (x^3 + x + 1)dx = \left(\frac{x^4}{4} + c_1\right) + \left(\frac{x^2}{2} + c_2\right) + (x + c_3)$$

$$= \frac{x^4}{4} + \frac{x^2}{2} + x + c$$

在最后的答案中,我们已将三个附下标的常数归并成一个常数 $c$。

作为一种通用的做法,积分过程中出现的所有相加的任意积分常数,都可以在最终答案中归并成一个任意常数。

**例 7** 求 $\int \left(2e^{2x} + \frac{14x}{7x^2 + 5}\right)dx$。

根据法则 IV,我们可以对被积函数中两个相加的项分别积分,然后加和其结果。因为 $2e^{2x}$ 项在积分法则 IIa 中的形式为 $f'(x)e^{f(x)}$,其中 $f(x) = 2x$,所以积分为 $e^{2x} + c_1$,类似地,另一项,$14x/$
442 $(7x^2 + 5)$,取形式 $f'(x)/f(x)$,有 $f(x) = 7x^2 + 5 > 0$。因此,由法则 IIIa,积分为 $\ln(7x^2 + 5) + c_2$。因此,可以写出

$$\int \left(2e^{2x} + \frac{14x}{7x^2 + 5}\right)dx = e^{2x} + \ln(7x^2 + 5) + c$$

其中,我们已将 $c_1$ 和 $c_2$ 合并为一个任意常数 $c$。

**法则 V(倍数的积分)** 一个常数 $k$ 与被积函数的积的积分,等于常数 $k$ 与被积函数积分的乘积。以符号表示,即

$$\int kf(x)dx = k\int f(x)dx$$

从运算角度看，此法则意味着乘积常数可以从积分符号中“提取”出去。（但要注意，含变量的项不能像这样提取出去。）为证明此法则，我们回忆一下，$k$ 乘以 $f(x)$ 仅意味着使 $f(x)$ 相加 $k$ 次，因而由法则 IV，有

$$\int kf(x)dx = \int \underbrace{[f(x)+f(x)+\cdots+f(x)]dx}_{k\text{项}}$$

$$= \underbrace{\int f(x)dx + \int f(x)dx + \cdots + \int f(x)dx}_{k\text{项}}$$

$$= k\int f(x)dx$$

**例 8** 求 $\int -f(x)dx$，这里 $k=-1$，因此，

$$\int -f(x)dx = -\int f(x)dx$$

此即函数负数的积分，等于函数积分的负数。

**例 9** 求 $\int 2x^2dx$。提取出 2 并应用法则 I，有

$$\int 2x^2dx = 2\int x^2dx = 2\left(\frac{x^3}{3}+c_1\right) = \frac{2}{3}x^3 + c$$

**例 10** 求 $\int 3x^2dx$。在本例中，提取乘积常数后得到

$$\int 3x^2dx = 3\int x^2dx = 3\left(\frac{x^3}{3}+c_1\right) = x^3 + c$$

注意，与上面一个例子相反，最终答案中的项 $x^3$ 不带有任何分数。这个简洁结果的出现是因为被积函数中的乘积常数恰好等于2（函数的幂）加 1。参考幂函数积分法则（法则 I），我们看到乘数

常数$(n+1)$在此情况下恰好消掉分数$1/(n+1)$，因此得到答案$(x^{n+1}+c)$。

一般而言，只要我们有被积函数$(n+1)x^n$，便没有必要提取常数$(n+1)$，然后再积分$x^n$，只需写出答案$x^{n+1}+c$即可。

**例11** 求$\int\left(5e^x - x^{-2} + \frac{3}{x}\right)dx$，$(x \neq 0)$。这个例子同时说明法则IV和V；实际上，它也说明了前三个法则：

$$\int\left(5e^x - \frac{1}{x^2} + \frac{3}{x}\right)dx = 5\int e^x dx - \int x^{-2} dx + 3\int \frac{1}{x} dx$$

[由法则IV和V]

$$= (5e^x + c_1) - \left(\frac{x^{-1}}{-1} + c_2\right) + (3\ln|x| + c_3)$$

$$= 5e^x + \frac{1}{x} + 3\ln|x| + c$$

此结果的正确性也可以通过微分来验证。

## 涉及代换的法则

现在，我们来介绍两个在适当情况下，可以通过代换原积分变量，从而使积分过程得以简化的法则。只要新引进的积分变量使积分过程比原来更简单，那么，这些法则就是有用的。

**法则VI(代换法则)** $f(u)(du/dx)$对变量$x$的积分，是$f(u)$对变量$u$的积分：

$$\int f(u)\frac{du}{dx}dx = \int f(u)du = F(u) + c$$

其中,运算$\int du$已经代替了运算$\int dx$。

此法则对应于积分学中的链式法则,也可以通过链式法则本身来加以证明。给定函数$F(u)$,其中$u=u(x)$,链式法则表明

$$\frac{d}{dx}F(u)=\frac{d}{du}F(u)\frac{du}{dx}=F'(u)\frac{du}{dx}=f(u)\frac{du}{dx}$$

因为$f(u)(du/dx)$是$F(u)$的导数,由(13.3)可知前者的积分(反导数)必为

$$\int f(u)\frac{du}{dx}dx=F(u)+c$$

读者也许会注意到,这个结果实际上也可以通过消去左边的两个 444
$dx$表达式而得到。

**例 12** 求$\int 2x(x^2+1)dx$。此题答案可以通过先乘开被积函数而得到:

$$\int 2x(x^2+1)dx=\int(2x^3+2x)dx=\frac{x^4}{2}+x^2+c$$

但现在,我们还是运用代换法则来解它。令$u=x^2+1$,则$du/dx=2x$,或$dx=du/2x$。以$du/2x$代换$dx$,得到

$$\int 2x(x^2+1)dx=\int 2xu\frac{du}{2x}=\int u\,du=\frac{u^2}{2}+c_1$$

$$=\frac{1}{2}(x^4+2x^2+1)+c_1=\frac{1}{2}x^4+x^2+c$$

其中$c=1/2+c_1$。以$du/dx$代换$2x$,也可以得到同样的结果。

**例 13** 求$\int 6x^2(x^3+2)^{99}dx$。本例中的被积函数很难乘开,因而现在是展示代换法则有效性的极好机会。令$u=x^3+2$,则$du/dx=3x^2$,从而

$$\int 6x^2(x^3+2)^{99}dx = \int \left(2\frac{du}{dx}\right)u^{99}dx = \int 2u^{99}du$$

$$= \frac{2}{100}u^{100} + c = \frac{1}{50}(x^3+2)^{100} + c$$

**例 14** 求 $\int 8e^{2x+3}dx$。令 $u = 2x+3$，那么，$du/dx = 2$，或 $dx = du/2$，因此

$$\int 8e^{2x+3}dx = \int 8e^u \frac{du}{2} = 4\int e^u du = 4e^u + c = 4e^{2x+3} + c$$

正如这些例子所表明的那样，只要我们审慎选择函数 $u = u(x)$，将被积函数（$x$ 的函数）表示成 $f(u)$（$u$ 的函数）和 $du/dx$（我们选择的 $u$ 函数的导数）的积。那么，这个法则是很有用的，而且，像后两个例子所表明的那样，当原被积函数被变换成 $f(u)(du/dx)$ 与常数的乘积时，还可以运用此法则。因为常数因子可以从积分符号中提取出去，留下形式为 $f(u)(du/dx)$ 的被积函数，它正是符合代换法则的要求。但是，当变量代换得到变量乘以 $f(u)(du/dx)$，比如，$x$ 乘 $f(u)(du/dx)$ 时，就不能提出乘积因子，应用代换法则。事实上，不存在一般的公式可以将两个函数乘积用两个函数各自的积分来表示；也没有一般的公式将两个函数商的积分，用两个函数单独的积分来表示。这就是为什么从总体上看积分较微分更困难一些，以及为什么对于复杂的被积函数，在现成的积分表中查找答案较之自己独自进行计算要方便一些的原因。

445

**法则 VII（分部积分）：** $v$ 对 $u$ 的积分等于 $uv$ 减去 $u$ 对 $v$ 的积分：

$$\int v\, du = uv - \int u\, dv$$

此法则的实质是以运算$\int dv$代替运算$\int du$。

这个结果的合理性是相对易于说明的。首先，由微分的乘积法则可知：

$$d(uv) = v\,du + u\,dv$$

若我们对方程两边积分(即对每一微分积分)，可得到一个新方程

$$\int d(uv) = \int v\,du + \int u\,dv$$

或
$$uv = \int v\,du + \int u\,dv$$

[方程左边不需加上常数(为什么?)]

则，在两边减去$\int u dv$，上述结论得证。

**例 15** 求$\int x(x+1)^{1/2}$。本例有别于例 12 和例 13，不能利用法则 VI 的代换法(为什么?)。但我们可以将给定积分视为$\int v du$的形式，并应用法则 VII，为此，令 $v=x$，这意味着$dv=dx$，且可令 $u=2/3(x+1)^{3/2}$，从而 $du=(x+1)^{1/2}$。则我们可以求得积分

$$\int x(x+1)^{1/2}dx = \int v\,du = uv - \int u\,dv$$

$$= \frac{2}{3}(x+1)^{3/2}x - \int \frac{2}{3}(x+1)^{3/2}dx$$

$$= \frac{2}{3}(x+1)^{3/2}x - \frac{4}{15}(x+1)^{5/2} + c$$

**例 16** 求$\int \ln x\,dx$，$(x>0)$。我们这里不能应用对数法则，因为此法则所涉及的是被积函数 $1/x$，而非 $\ln x$。我们也不能应用法则 VI。但我们若令 $v=\ln x$，便有 $dv=(1/x)dx$，再令 $u=$ 446

$x$，从而有 $du = dx$，则积分过程如下

$$\int \ln x\, dx = \int v\, du = uv - \int u\, dv$$

$$= x \ln x - \int dx = x \ln x - x + c = x(\ln x - 1) + c$$

**例 17** 求 $\int xe^x\, dx$。在本例中，仅令 $v = x$，$u = e^x$，从而有 $dv = dx$，$du = e^x\, dx$。应用法则 VII，则有

$$\int xe^x\, dx = \int v\, du = uv - \int u\, dv$$

$$= e^x x - \int e^x dx = e^x x - e^x + c = e^x(x - 1) + c$$

同上面所有例子一样，此结果的正确性完全可通过微分来加以检验。

## 练习 13.2

---

1 求下列积分：

(a) $\int 16x^{-3} dx$ $(x \neq 0)$

(b) $\int 9x^8 dx$

(c) $\int (x^5 - 3x) dx$

(d) $\int 2e^{-2x} dx$

(e) $\int \frac{4x}{x^2 + 1} dx$

(f) $\int (2ax + b)(ax^2 + bx)^7 dx$

2 求下列积分：

(a) $\int 13e^x dx$

(b) $\int \left(3e^x + \frac{4}{x}\right) dx$ $(x > 0)$

(c) $\int \left(5e^x + \frac{3}{x^2}\right) dx$ $(x \neq 0)$

(d) $\int 3e^{-(2x+7)} dx$

(e) $\int 4xe^{x^2+3} dx$

(f) $\int xe^{x^2+9} dx$

3　求

(a) $\int \frac{3dx}{x}$ $(x \neq 0)$　　(c) $\int \frac{2x}{x^2+3}dx$

(b) $\int \frac{dx}{x-2}$ $(x \neq 2)$　　(d) $\int \frac{x}{3x^2+5}dx$

4　求

(a) $\int (x+3)(x+1)^{1/2}dx$　　(b) $\int x\ln x dx$ $(x>0)$

5　给定 $n$ 个常数 $k_i(i=1,2,\cdots,n)$ 和 $n$ 个函数 $f_i(x)$，根据法则 IV 和 V 推导出

$$\int \sum_{i=1}^{n} k_i f_i(x)dx = \sum_{i=1}^{n} k_i \int f_i(x)dx$$

---

## 13.3　定积分



### 定积分的含义

上一节所列出的所有积分均属于不定积分：每个积分都是变量的函数，因而没有确定的数值。现在，对于连续函数 $f(x)$ 的已知不定积分，

$$\int f(x)dx = F(x) + c$$

若我们选择 $x$ 定义域中的两个值，比如 $a$ 和 $b(a<b)$，依次将其代入方程右边，并形成差

$$[F(b)+c] - [F(a)+c] = F(b) - F(a)$$

从而得到一个不再包含变量 $x$ 及任意常数 $c$ 的具体数值。此值称作 $f(x)$ 从 $a$ 至 $b$ 的定积分，其中 $a$ 为积分下限，$b$ 为积分上限。

为标示出积分上下限，我们将积分符号修改成这种形式 $\int_a^b$。

则可用符号将定积分的计算步骤表示如下：

$$\int_a^b f(x)dx = F(x)\Big]_a^b = F(b) - F(a) \qquad (13.6)$$

其中符号 $\Big]_a^b$（也可以写成 $\Big|_a^b$ 或 $\Big[\cdots\Big]_a^b$）表示用 $a$ 和 $b$ 依次代替积分结果中的 $x$ 以得到 $F(b)$ 和 $F(a)$，然后再取其差，如(13.6)右边所示。第一步，我们必须先求出不定积分，但可以省略常数 $c$，因为在取差的过程中可将其消掉。

**例 1** 计算 $\int_1^5 3x^2dx$。因为不定积分为 $x^3 + c$，所以定积分的值为

$$\int_1^5 3x^2dx = x^3\Big]_1^5 = (5)^3 - (1)^3 = 125 - 1 = 124$$

**例 2** 计算 $\int_a^b ke^x dx$。这里，积分的上下限是以符号形式给出的，因而，积分的结果也以符号表示：

$$\int_a^b ke^x dx = ke^x\Big]_a^b = k(e^b - e^a)$$

448 **例 3** 计算 $\int_0^4\left(\frac{1}{1+x} + 2x\right)dx$，$(x \neq -1)$。不定积分是 $\ln|1+x| + x^2 + c$；因此，答案是

$$\int_0^4\left(\frac{1}{1+x} + 2x\right)dx = \Big[\ln|1+x| + x^2\Big]_0^4$$

$$= (\ln 5 + 16) - (\ln 1 + 0)$$

$$= \ln 5 + 16 \ [\text{因为} \ln 1 = 0]$$

重要的是，要认识到积分的上下限 $a$ 和 $b$ 均是变量 $x$ 定义域中的值。如果在积分时运用变量代换法(法则 VI 和法则 VII)，并引入变量 $u$，就不能把 $a$ 和 $b$ 视为 $u$ 的积分限。下面的例子将说明这一点。

**例 4** 计算 $\int_1^2 (2x^3-1)^2(6x^2)dx$。令 $u=2x^3-1$, $du/dx=6x^2$,或 $du=6x^2dx$。注意,当 $x=1$, $u$ 将等于 1,但当 $x=2$ 时,$u$ 将为 15。换言之,变量 $u$ 的积分限将为 1(下限)和 15(上限)。因此,将给定的积分以 $u$ 的形式重写,不会得到 $\int_1^2 u^2du$,而是

$$\int_1^{15} u^2du = \frac{1}{3}u^3\Big]_1^{15} = \frac{1}{3}(15^3-1^3) = 1124\frac{2}{3}$$

此外,我们还可以将 $u$ 变换回 $x$,然后运用原来的积分限 1 和 2 以得到同样的答案:

$$\left[\frac{1}{3}u^3\right]_{u=1}^{u=15} = \left[\frac{1}{3}(2x^3-1)^3\right]_{x=1}^{x=2} = \frac{1}{3}(15^3-1^3) = 1124\frac{2}{3}$$

## 作为曲线下面积的定积分

每个积分均有一个确定的值。在几何上,此值可以解释为一条给定曲线下的特定面积。

连续函数 $y=f(x)$的图形绘在图 13.1 中。若我们试图度量由曲线和横轴围起来的,位于定义域中两点 $a$ 和 $b$ 的面积 $A$(图中阴影部分),我们可以采纳如下步骤:首先,将区间$[a,b]$划分成 $n$ 个子区间(长度可以不相等)。图 $a$ 中绘出了四个子区间(即 $n=4$),第一个是$[x_1,x_2]$,最后一个是$[x_4,x_5]$。因为其中的每一个都表示 $x$ 的变化,所以可以分别将其视为 $\Delta x_1,\cdots,\Delta x_4$。现在,我们在这些区间上作出四个矩形块,使得每个矩形的高度等于函数在该矩形中可达到的最大值(这里恰好等于每个矩形的左边界)。因此,第一个矩形的高度为 $f(x_1)$,宽度为 $\Delta x_1$;一般地,第 $i$ 个矩形块的高为$f(x_i)$,宽为 $\Delta x_i$。这组矩形的总面积 $A^*$ 等于和

449

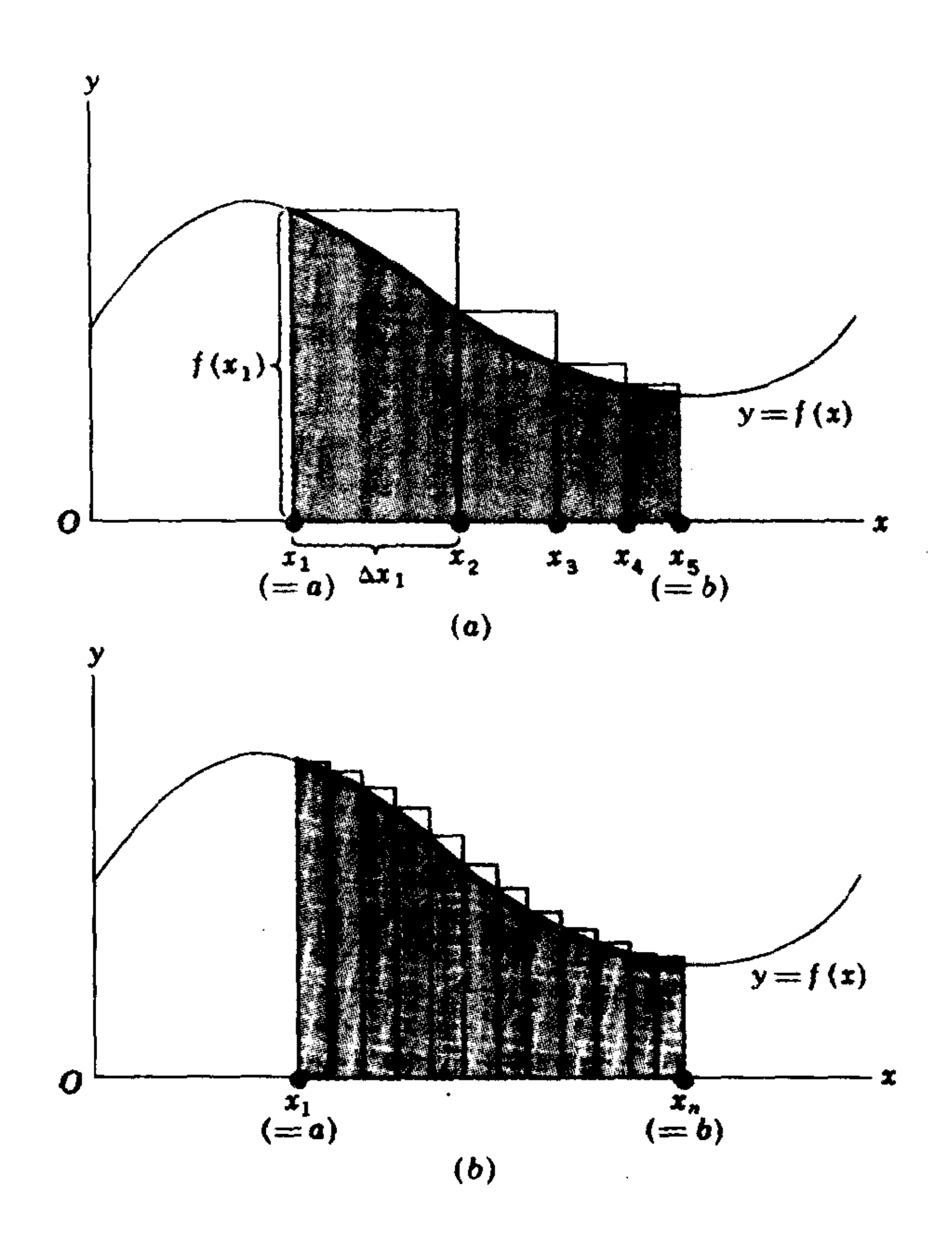



图 13.1

$$A^* = \sum_{i=1}^{n} f(x_i)\Delta x_i (n = 4 \text{ 在图 } 13.1a)$$

它显然不是我们要求的曲线下的面积,只是曲线下面积的一个大致的近似。

$A$ 的真实值与 $A^*$ 的差别在于矩形块中未加阴影的部分;这部分使 $A^*$ 大于 $A$。但如果可以减缩阴影面积并使其趋于零,则相应地,近似值 $A^*$ 也可以趋于真实值 $A$。当我们将区间[$a$,$b$]

分割得越来越细,从而从 $n$ 无限增大,$\Delta x_i$ 无限缩小时,这个结果就可以成为现实。这样,上述的那种矩形块将变得越来越细长,曲线外的长出部分就越来越小;从图 $b$ 中我们可以见到这一点。这种使矩形"苗条化"的运算,在极限情况下成为 450

$$\lim_{n\to\infty}\sum_{i=1}^{n}f(x_i)\Delta x_i = \lim_{n\to\infty}A^* = \text{面积 } A \qquad (13.7)$$

如果此极限存在的话,而在本例中,极限确实存在。此方程实际上就是曲线下面积的正式定义。

(13.7)中的和表达式 $\sum_{i=1}^{n}f(x_i)\Delta x_i$ 与定积分表达式 $\int_a^b f(x)dx$ 有某种相似之处。事实上,后者是以前者为基础的。当变化 $\Delta x_i$ 无穷小时,我们可以用 $dx_i$ 来代替它。而且,下标 $i$ 可以舍去,因为每个无穷小变化都可以同样符号 $dx$ 表示。因此,我们可以将 $f(x_i)\Delta x_i$ 重写成 $f(x)dx$。求和符号怎么办呢? 求和符号 $\sum_{i=1}^{n}$ 表示有限项之和。含 $n\to\infty$,并取和的极限时,这个经常使用的符号就有些累赘了。因此需要一个更简单的替代符号。这个替代符号即 $\int_a^b$,其中加长的 $S$ 符号也表示和,$a$ 和 $b$(相当于 $i=1$ 和 $i=n$)用于设定此和的下限和上限。总之,定积分是和的极限表达式(13.7)的简写。即

$$\int_a^b f(x)dx \equiv \lim_{n\to\infty}\sum_{i=1}^{n}f(x_i)\Delta x_i = \text{面积 } A$$

因此,上面所说的定积分(称作黎曼积分)现在则有了面积的含义及和的含义,因为 $\int_a^b$ 是与离散的 $\sum_{i=1}^{n}$ 概念对应的连续的概念。

在图 13.1 中，我们力图通过细分区间$[a,b]$来系统降低过剩近似面积 $A^*$，使之近似表示面积 $A$。由此而产生的矩形块面积的和的极限，称作上积分，它是从曲线上方趋近于 $A$ 的近似值。我们也可以从曲线下方作出内接而不是伸出曲线外的矩形块来近似面积 $A$(见练习 13.3-3)。这组新的矩形块的总面积 $A^{**}$ 会低估面积 $A$，但随着区间$[a,b]$越来越细分，我们又可求得 $\lim_{n\to\infty} A^{**} = A$。这个矩形块面积和的极限称作下积分。当且仅当上积分和下积分的值相等时，黎曼积分 $\int_a^b f(x)dx$ 才有定义，而函数 $f(x)$ 则称作黎曼可积的。存在一些定理来设定函数 $f(x)$ 可积的条件。按照微积分的基本定理，如果函数在区间$[a,b]$连续，则它在
451 该区间中便可积。因此，只要我们讨论的是连续函数，就不必担心是否可积的问题。

还有一点要注意。虽然图 13.1 中的面积 $A$ 恰好完全位于曲线$y=f(x)$的递减部分，但对于向上倾斜的曲线下的面积，定积分的概念也完全适用。实际上，曲线的向上和向下倾斜可以同时存在，例如，我们可以 $\int_0^b f(x)dx$ 来计算图 13.1 中曲线下 $Ob$ 间的面积。

注意，如果我们以定积分 $\int_a^b f(x)dx$ 计算图 13.2 中的面积 $B$，结果会是负值，因为该面积中所包含的矩形块的高度均是负值。这就产生了负面积的概念，它是已知曲线上方，$x$ 轴下方的面积。在我们感兴趣的是面积的数值，而非面积的代数值的情况下，我们就取相关定积分的绝对值。另一方面，面积 $C = \int_c^d f(x)dx$ 是正值，尽管它位于 $x$ 轴的负值区域，因为当由 $c$ 移至$d$ 时，每个

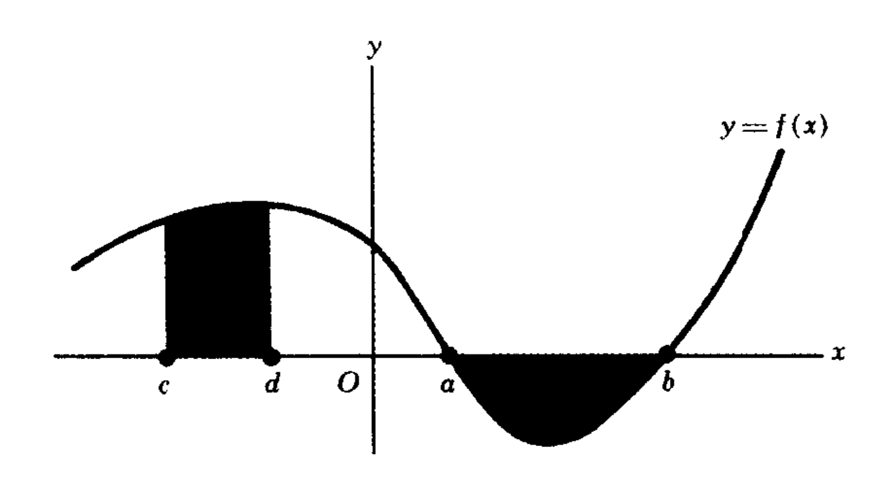



图 13.2

矩形块的高和宽均是正值。由此可知,通过改变移动方向互换积分上下限,将改变 $\Delta x_i$ 的符号和定积分的符号,这个含义是很明显的。应用于面积 $B$,我们看到定积分 $\int_b^a f(x)dx$ (由 $b$ 至 $a$)将给出面积 $B$ 的度量。它将度量这个面积的数值。

## 定积分的性质

上一小节的讨论,使我们得到如下定积分性质。

**性质 I** 定积分限的互换,使定积分的符号改变: 452

$$\int_b^a f(x)dx = -\int_a^b f(x)dx$$

此性质证明如下:

$$\int_b^a f(x)dx = F(a) - F(b) = -[F(b) - F(a)] = -\int_a^b f(x)dx$$

定积分还有其它重要性质。

**性质 II** 当一个定积分的上限与下限相等时,该定积分的值为零。

$$\int_a^a f(x)dx = F(a) - F(a) = 0$$

按照"面积"的解释,这意味着定义域中任意一点以上,曲线以

下的面积等于零。这也应当如此，因为在 $x$ 轴上某点的上方，只能绘出一条（一维的）直线，不能绘出（二维的）面积。

**性质 III** 一个定积分可以表示成如下有限个子定积分的和：

$$\int_a^d f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx + \int_c^d f(x)dx$$

$$(a < b < c < d)$$

在此方程中只给出三个子积分，但该性质对 $n$ 个子积分的情况同样成立。有时也称此性质为可加性。

根据面积的解释，这意味着 $x$ 轴上面区间 $[a,d]$ 的面积可以通过相加子区间 $\{[a,b],[b,c],[c,d]\}$ 的面积得到。注意，因为这里讨论的是闭集，所以边界点 $b$ 和 $c$ 每次都被包括在两个面积中。这是否涉及双重计算问题？确实如此。但幸运的是，这不会造成任何不利影响，因为根据性质 II，单一点上面的面积为零，因此双重计算对结果不会产生任何影响。但是，无庸置疑，任何区间的双重计算是绝不允许的。

前面，我们曾提到任何连续函数均是黎曼可积的。现在，根据性质 III，我们还可以求得某些不连续函数的定积分（面积）。考察图 13.3$a$ 中的分段函数。尽管此函数在区间 $[a,c]$ 中的点 $b$ 不连续，我们可以由如下和

$$\int_a^b f(x)dx + \int_b^c f(x)dx$$

求得阴影面积。这也可以应用于图 $b$ 中的曲线。

453 **性质 IV**

$$\int_a^b -f(x)dx = -\int_a^b f(x)dx$$

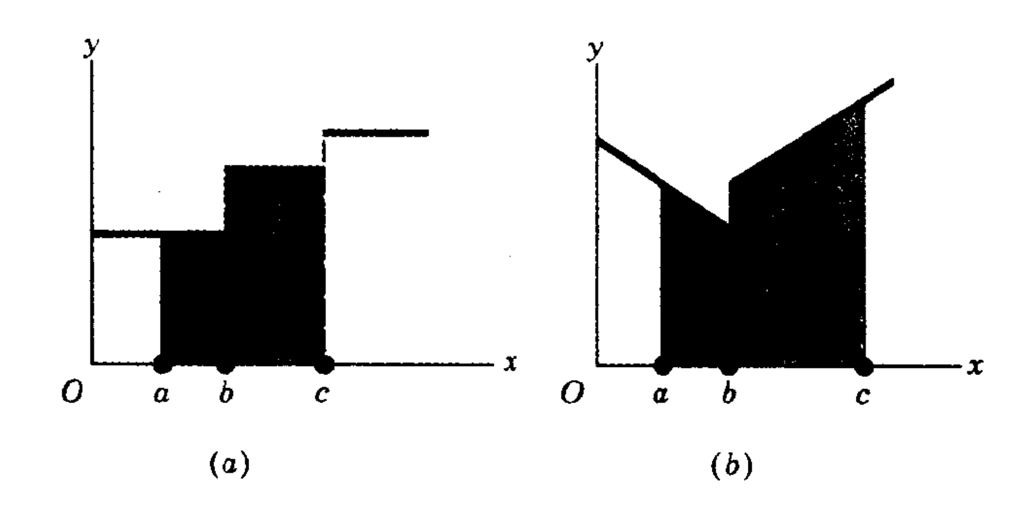



图 13.3

**性质 V**

$$\int_a^b kf(x)dx = k\int_a^b f(x)dx$$

**性质 VI**

$$\int_a^b [f(x) + g(x)]dx = \int_a^b f(x)dx + \int_a^b g(x)dx$$

**性质 VII （分步积分）** 给定 $u(x)$和 $v(x)$,

$$\int_{x=a}^{x=b} v\,du = uv\Big|_{x=a}^{x=b} - \int_{x=a}^{x=b} u\,dv$$

这四个性质均借自不定积分法则,无需进一步解释。

## 对不定积分的进一步考察

通过给不定积分加上两个积分限的方式,我们引入了定积分。现在我们已知道了定积分的含义,我们再来考察如何从定积分的含义反推不定积分的含义。

假设我们不是把积分上限固定在 $b$,而是让它为一个变量,称为 $x$。那么将取这种形式

$$\int_a^x f(x)dx = F(x) - F(a)$$

它现在是 $x$ 的函数，表示 $f(x)$曲线下的可变面积。但因为右边最后一项是一个常数，因此这个积分必定是原函数 $f(x)$亲系中的一员，在前面，我们将其表示为 $F(x)+c$。若令 $c=-F(a)$，则上述积分恰好变成不定积分 $\int f(x)dx$。

454

据此，我们可以认为符号 $\int$ 与符号 $\int_a^x$ 具有相同含义，只要我们将 $\int_a^x$ 中的积分下限理解为以方程 $c=-F(a)$表示的积分常数。

## 练习 13.3

---

1 计算如下定积分

(a) $\int_1^3 \frac{1}{2}x^2dx$　　(d) $\int_2^4(x^3-6x^2)dx$

(b) $\int_0^1 x(x^2+6)dx$　　(e) $\int_{-1}^1(ax^2+bx+c)dx$

(c) $\int_1^3 3\sqrt{x}dx$　　(f) $\int_4^2 x^2\left(\frac{1}{3}x^3+1\right)dx$

2 计算如下定积分

(a) $\int_1^2 e^{-2x}dx$　　(c) $\int_2^3(e^{2x}+e^x)dx$

(b) $\int_{-1}^{e-2}\frac{dx}{x+2}$　　(d) $\int_e^6\left(\frac{1}{x}+\frac{1}{1+x}\right)dx$

3 在图 13.1$a$ 中，取在每个子区间中函数的最小值作为矩形块的高度，即取 $f(x_2)$而非 $f(x_1)$作为第一个矩形块的高，但仍以 $\Delta x_i$ 为矩形块的宽；对其它矩形块也作类似处理。

($a$) 写出新的矩形的总面积 $A^{**}$ 的求和表达式。

($b$) $A^{**}$ 是高估还是低估了所求的面积 $A$。

($c$)如果进一步细分区间[$a$,$b$],$A^{**}$是趋近还是偏离 $A$? [提示:用图形描述。]

($d$)在极限情况下,当子区间数 n 趋于无穷大时,近似值 $A^{**}$ 是否像 $A^*$ 那样,趋近于真实值 $A$?

($e$)根据上述讨论,关于图中函数 $f(x)$ 的黎曼可积性,可以得出什么结论?

4 定积分 $\int_a^b f(x)dx$ 表示某曲线下的面积,这条曲线是指被积函数 $f(x)$,还是原函数 $F(x)$? 若我们绘出函数 $F(x)$ 的图形,我们如何在这个图形上表示出上述定积分? 是一个面积、一个线段、还是一个点?

5 证明常数 $c$ 可以等价地表示为

$$(a)\ c \equiv \int_0^b \frac{c}{b}dx \qquad (b)\ c \equiv \int_0^b 1dt$$

---

## 13.4 广义积分

某些积分被称作"广义积分"。我们将简要讨论其中的两种类型。

### 无穷极限积分

455

当我们有如下形式的定积分

$$\int_a^{\infty} f(x)dx \text{ 和 } \int_{-\infty}^{b} f(x)dx$$

其中有一个积分限是无穷大时,则我们称此积分为**广义积分**。在这种情况下,不能像如下方式分别计算积分

$$F(\infty) - F(a) \qquad \text{和} \qquad F(b) - F(-\infty)$$

因为 $\infty$ 不是一个数,因此,它不能替代函数 $F(x)$ 中的 $x$。所以,我们必须再一次借助于极限的概念。

上面列出的第一个积分可以定义为另一个积分当其积分上限趋于无穷大时的极限，即

$$\int_{a}^{\infty} f(x)dx \equiv \lim_{b\to\infty}\int_{a}^{b} f(x)dx \tag{13.8}$$

若此极限存在，则称此广义积分收敛，且此取极值的过程会产生积分值。若极限不存在，则称此广义积分发散，因而是无意义的。同理，我们可以定义

$$\int_{-\infty}^{b} f(x)dx \equiv \lim_{a\to-\infty}\int_{a}^{b} f(x)dx \tag{13.8'}$$

它也有同样的收敛与发散的判别标准。

**例 1** 计算广义积分 $\int_{1}^{\infty}\frac{dx}{x^2}$。先要注意

$$\int_{1}^{b}\frac{dx}{x^2} = \frac{-1}{x}\Big]_{1}^{b} = \frac{-1}{b} + 1$$

因此，按照(13.8)，所求积分为

$$\int_{1}^{\infty}\frac{dx}{x^2} = \lim_{b\to\infty}\int_{1}^{b}\frac{dx}{x^2} = \lim_{b\to\infty}\left(\frac{-1}{b} + 1\right) = 1$$

此广义积分的确收敛，且其值为 1。

因为极限表达式写起来有些麻烦，有些人宁愿省略"lim"符号，仅写成

$$\int_{1}^{\infty}\frac{dx}{x^2} = \frac{-1}{x}\Big]_{1}^{\infty} = 0 + 1 = 1$$

即便写成这种形式，广义积分仍要解释成极限概念。

从图形上看，广义积分仍具有面积的含义。但因在此情况下积分上限可取不断递增的值，所以右边界必然向东无限延伸，如图 13.4*a* 所示。尽管如此，我们仍可视此面积具有确定的(极限)值 1。

456

**例2** 计算$\int_1^{\infty}\frac{dx}{x}$。同前面一样，我们先求

$$\int_1^b \frac{dx}{x} = \ln x\Big]_1^b = \ln b - \ln 1 = \ln b$$

当我们令 $b\to\infty$时，由(10.16′)我们有 $\ln b\to\infty$。因此，这个广义积分是发散的。

图 13.4$b$ 给出了函数 $1/x$ 的图形，以及对应于给定积分的面积。这次，右边界向右无限延伸，将使面积无限增大，尽管此积分的图形非常类似于图 $a$ 的图形。

如果积分的上、下限均为无穷大，又将如何呢？(13.8)和(13.8′)的直接推广将给出如下定义

$$\int_{-\infty}^{\infty} f(x)dx = \lim_{\substack{b\to+\infty\\ a\to-\infty}} \int_a^b f(x)dx \qquad (13.8'')$$

同样，当且仅当此广义积分的极限存在时，则称此广义积分是收敛的。

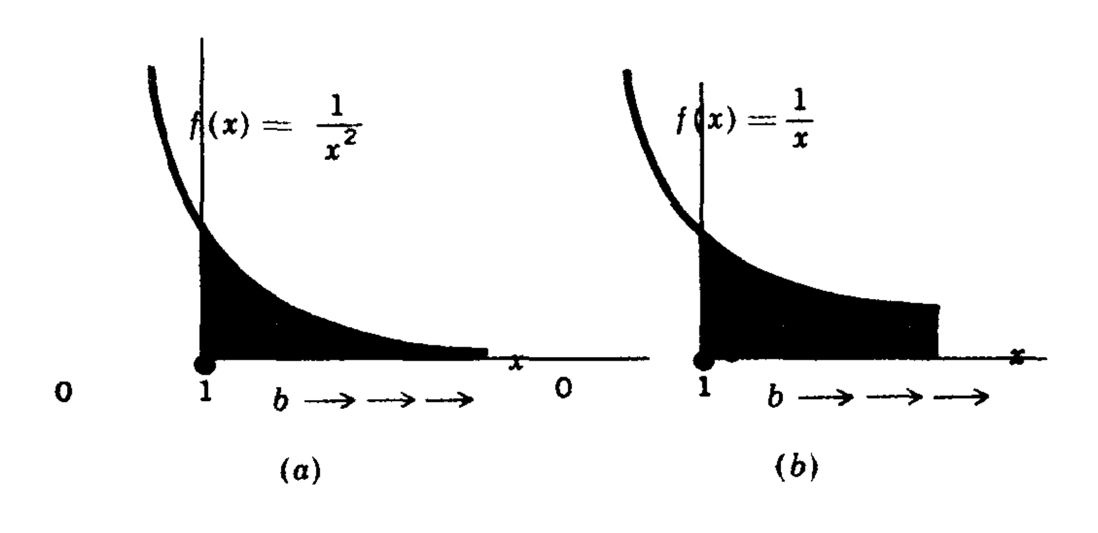



图 13.4

## 无穷被积函数

既使具有确定的积分限，如果被积函数在积分区间[$a$，$b$]中

的某处变得无穷大,那么,这个积分仍是广义积分。要计算这样的积分,我们仍需依靠极限概念。

**例 3** 计算 $\int_0^1 \frac{1}{x}dx$ 。因为如图 13.4$b$ 所示,此被积函数在积分下限为无穷大(当 $x \to 0^+$ 时,$1/x \to \infty$),所以,此积分是广义积分。因此,我们先求积分

457

$$\int_a^1 \frac{1}{x}dx = \ln x\Big]_a^1 = \ln 1 - \ln a = -\ln a \ [\text{由于 } a > 0]$$

然后计算当 $a \to 0^+$ 时的极限:

$$\int_0^1 \frac{1}{x}dx \equiv \lim_{a \to 0^+} \int_a^1 \frac{1}{x}\mathrm{d}x = \lim_{a \to 0^+} (-\ln a)$$

因为此极限不存在(当 $a \to 0^+$ 时,$\ln a \to -\infty$),所以,此积分是发散的。

**例 4** 计算 $\int_0^9 x^{-1/2}dx$。当 $x \to 0^+$,被积函数 $1/\sqrt{x}$ 为无穷大,此积分为广义积分。同样,我们先求

$$\int_a^9 x^{-1/2}dx = 2x^{1/2}\Big]_a^9 = 6 - 2\sqrt{a}$$

当 $a \to 0^+$ 时,此式的极限是 $6-0=6$。因此,该给定积分收敛于 6。

被积函数在积分上限为无穷大的情形也完全类似。但是,被积函数在开区间$(a,b)$而非在 $a$ 点或 $b$ 点达到无穷大,则是完全不同的命题。在这种意外情况下,必须利用定积分的可加性,并先把给定区间分割成子区间。假设当 $x \to p$ 时 $f(x) \to \infty$,其中 $p$ 是区间$(a,b)$中的点,则由可加性,我们有

$$\int_a^b f(x)dx = \int_a^p f(x)dx + \int_p^b f(x)dx$$

当且仅当每个子区间有极限时,上式左边的给定区间才可以视为

收敛的。

**例 5**　计算 $\int_{-1}^{1} \frac{1}{x^3} dx$。当 $x \to 0$ 时，被积函数趋于无穷大。因此，必须将给定区间写成和

$$\int_{-1}^{1} x^{-3} dx = \int_{-1}^{0} x^{-3} dx + \int_{0}^{1} x^{-3} dx \text{（即，} \equiv I_1 + I_2\text{）}$$

区间 $I_1$ 是发散的，因为

$$\lim_{b \to 0^-} \int_{-1}^{b} x^{-3} dx = \lim_{b \to 0^-} \left[ \frac{-1}{2} x^{-2} \right]_{-1}^{b} = \lim_{b \to 0^-} \left( -\frac{1}{2b^2} + \frac{1}{2} \right) = -\infty$$

因此，无须计算 $I_2$，我们便可以马上得出结论：此给定区间是发散的。

## 练习 13.4

458

---

1　考察练习 13.3－1 和 13.3－2 给出的定积分，确定其中的积分是否是广义积分。如果是，属于哪类广义积分？

2　下列积分哪个是广义积分？为什么？

(a) $\int_0^{\infty} e^{-rt} dt$　　(c) $\int_0^1 x^{-2/3} dx$　　(e) $\int_1^5 \frac{dx}{x-2}$

(b) $\int_2^3 x^4 dx$　　(d) $\int_{-\infty}^{0} e^{rt} dt$　　(f) $\int_{-3}^{4} 6 dx$

3　计算上题中所有的广义积分。

4　计算例 5 中的区间 $I_2$，证明它也是发散的。

5　(a)对于非负的 $t$，($c>0$)，绘出函数 $y = ce^{-t}$ 的图形，并绘出曲线下的阴影面积。

(b) 写出此面积的数学表达式，并确定它是否是有限的面积。

---

## 13.5 积分的经济应用

在经济分析中以各种不同方式应用积分。我们将在本节介绍一些简单应用，然后在下一节介绍其在多马增长模型中的应用。

### 从边际函数到总函数

给定一个总函数(比如总成本函数)，对其微分则会产生边际函数(比如边际成本函数)。由于积分过程与微分过程恰好相反，所以它使我们可以从已知的边际函数反推出总成本函数。

**例1** 如果厂商的边际成本($MC$)是产出的下述函数：$C'(Q)=2e^{0.2Q}$，若固定成本 $C_F=90$，求总成本函数 $C(Q)$。将 $C'(Q)$对 $Q$ 积分，我们求得

$$\int 2e^{0.2Q}dQ = 2\frac{1}{0.2}e^{0.2Q} + c = 10e^{0.2Q} + c \qquad (13.9)$$

此结果可以视为所求的 $C(Q)$函数，但有一点除外，即由于任意常数 $c$，此答案似乎是未确定的。幸运的是，信息 $C_F=90$ 可以作为确定常数 $c$ 的初始条件。当 $Q=0$ 时，总成本 $C$ 将仅含有 $C_F$。令(13.9)中的 $Q=0$，得到一个值 90，即 $10e^0+c=90$，这意味着 $c=90-10=80$。因此，总成本函数为

$$C(Q) = 10e^{0.2Q} + 80$$

459 注意，这里与(13.2)中的情况不同。(13.2)中的任意常数 $c$ 同变量 $H(0)$的初始值相同。在本例中，我们有 $c=80$，但 $C(0)\equiv C_F=90$，因而二者取不同的值。一般而言，不应假定任意常数 $c$ 总是等

于总函数的初始值。

**例2** 如果边际储蓄倾向($MPS$)是收入的如下函数:$S'(Y)=0.3-0.1Y^{-1/2}$,若当收入 $Y=81$ 时,总储蓄 $S=0$,求储蓄函数$S(Y)$。因为 $MPS$ 是 $S$ 函数的导数,现在的问题是求 $S'(Y)$的积分:

$$S(Y)=\int(0.3-0.1Y^{-1/2})dY=0.3Y-0.2Y^{1/2}+c$$

根据当 $Y=81$ 时,$S=0$,可以求出任意常数 $c$ 的具体值。尽管严格地讲,这并不是一个初始条件(与 $Y=0$ 不相关),但将此信息代入上述积分可以确定 $c$ 的值。因为

$$0=0.3(81)-0.2(9)+c\Rightarrow c=-22.5$$

所求的储蓄函数为

$$S(Y)=0.3Y-0.2Y^{1/2}-22.5$$

上述两个例子介绍的方法可直接推广至由已知边际函数求总函数(如总收益函数、总消费函数)的其它问题。还应重申的是,在这类问题中,答案(积分)的正确性,总可以通过微分来加以检验。

## 投资与资本形成

资本形成是增加给定资本存量的过程。将此过程视为一个连续过程,我们可以将资本存量表示成时间的函数 $K(t)$,并以导数 $dK/dt$ 表示资本形成率。[①] 但是,在时间 $t$ 的资本形成率与以 $I(t)$表示的净投资(流量)率相等。因此,资本存量 $K$ 和净投资 $I$

---

① 作为一个符号,变量对时间的导数通常也以在变量上置一点来表示,比如 $\dot{K}\equiv dK/dt$。在动态分析中,经常出现对时间的导数,使用这一更简洁的符号可以使符号表示更简明扼要。但是,这个符号只是一个小点,容易被忽略或放错位置。因此,在使用这种符号时应加倍小心。

通过如下两个方程联系起来：

$$\frac{dK}{dt} \equiv I(t)$$

和 $$K(t) = \int I(t)dt = \int \frac{dK}{dt}dt = \int dK$$

460 上面第一个方程是一个恒等式，它说明净投资与资本增加意义相同。因为 $I(t)$是 $K(t)$的导数，显然 $K(t)$是 $I(t)$的积分或反倒数，正如第二个方程所示。后一个方程中的被积函数的变换也是易于理解的：从 $I$ 转变为 $dK/dt$ 是根据定义，后一个变换则是消去两个相等的微分，即根据代换法则。

有时总投资的概念也与净投资一起在模型中使用。我们以 $I_g$ 表示总投资，$I$ 表示净投资，二者可通过方程

$$I_g = I + \delta K$$

联系起来。其中 $\delta$ 表示资本折旧率，$\delta K$ 表示重置投资率。

**例 3** 假设净投资流量以方程 $I(t) = 3t^{1/2}$表示，在时间 $t = 0$ 时的初始资本存量是 $K(0)$。何谓资本 $K$ 的时间路径？将 $I(t)$ 对 $t$ 积分，我们得到

$$K(t) = \int I(t)dt = \int 3t^{1/2}dt = 2t^{3/2} + c$$

其次，令最左边和最右边表达式中的 $t = 0$，求得 $K(0) = c$。因而，$K$ 的时间路径为

$$K(t) = 2t^{3/2} + K(0) \tag{13.10}$$

注意观察结果(13.10)和(13.2″)间的基本相似性。

当有人期望求某一时间区间的资本形成数量（而非资本 $K$ 的时间路径）时，就需使用定积分的概念。因为 $\int I(t)dt = K(t)$，我们可以写出定积分

$$\int_a^b I(t)dt = K(t)\Big]_a^b = K(b) - K(a)$$

来表示时间区间$[a,b]$的总资本积累。当然,它也可以表示 $I(t)$ 曲线下的面积。但应注意,在 $K(t)$函数的图形中,定积分表示一段垂直距离,更具体地说,表示两个垂直距离 $K(b)$与 $K(a)$之差。(参见练习 13.3－4。)

为更充分地理解 $K(t)$与 $I(t)$之间的不同,我们强调资本 $K$ 是一个存量的概念,而投资 $I$ 是一个流量的概念。因而,$K(t)$表示在每一时点存在的 $K$ 的数量,而 $I(t)$则给出每年或某一时期的净投资率,该投资率在该时期内是一致的。因此,为计算所进行的净投资数量(资本积累),我们必须首先设定所涉及的时期的长度。当我们将恒等式 $dK/dt \equiv I(t)$重写为 $dK \equiv I(t)dt$ 时,也可 461 以看到这一事实。$dK \equiv I(t)dt$ 表明,$K$ 的增量 $dK$ 不仅以流量变化率 $I(t)$为基础,而且以逝去的时间 $dt$ 为基础。正是由于设定表达式 $I(t)dt$ 中的时期的需要,才要进行定积分,并以它来表示 $I(t)$曲线[与 $K(t)$曲线相对]下的面积。

**例 4**　若净投资是一个不变流量 $I(t)=1000$(美元/年),那么,在一年内,即由 $t=0$ 至 $t=1$ 的总净投资(资本形成)是多少?显然,答案是 1000 美元;正式地,此结果可以如下方式得出:

$$\int_0^1 I(t)dt = \int_0^1 1000dt = 1000t\Big]_0^1 = 1000$$

读者可以验证,如果所涉及的年份是 $t=1$ 至 $t=2$,答案依然相同。

**例 5**　若 $I(t)=3t^{1/2}$(千美元/年),这是一个可变流量。那么,时期$[1,4]$,即第二、第三、第四年间的资本形成为多少?答案在于定积分

$$\int_1^4 3t^{1/2}dt = 2t^{3/2}\Big]_1^4 = 16 - 2 = 14$$

在上例的基础上,我们可以用定积分

$$\int_0^t I(t)dt = K(t)\Big]_0^t = K(t) - K(0)$$

表示对于投资率 $I(t)$,在时间区间$[0,t]$的资本积累数量。图 462 13.5描述了时期$[0,t_0]$的情况。从另一个角度看,上述方程产生了对于时间路径 $K(t)$的如下表达式

$$K(t) = K(0) + \int_0^t I(t)dt$$

在任意时间 $t$ 的$K$ 的数量,等于原始资本加上自那时起的总资本积累。

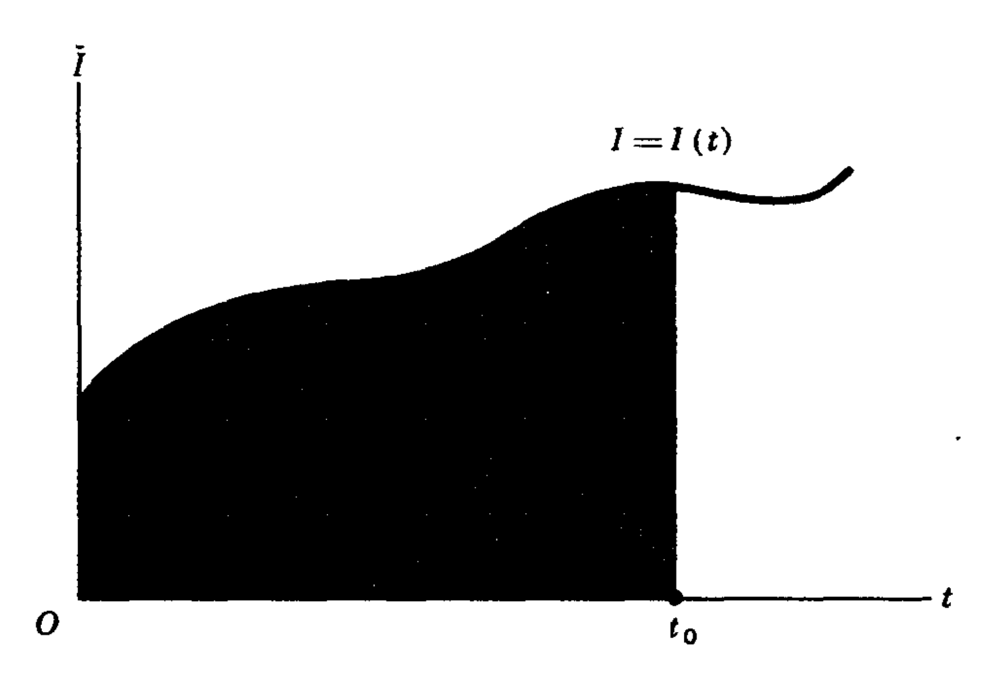



图 13.5

## 资金流量的现值

我们前面对贴现和现值问题的讨论,仅局限于单一的未来值

$V$ 的情况，得到的贴现公式是

$$A = V(1+i)^{-t} \qquad [离散的情况]$$

和

$$A = Ve^{-rt} \qquad [连续的情况]$$

现在，我们假设有一个未来值的流——在未来各个时间可获得的一系列收益，或在各个时间要支付的成本。那么，我们如何计算整个现金流的现值呢？

在离散情况下，若我们假设有三个在 $t$ 年末可获得的收益数字 $Rt(t=1,2,3)$，每年的利息率为 $i$，那么，$Rt$ 的现值分别为

$$R_1(1+i)^{-1} \quad R_2(1+i)^{-2} \quad R_3(1+i)^{-3}$$

由此得总现值为和

$$\Pi = \sum_{t=1}^{3} R_t(1+i)^{-t} \tag{13.11}$$

（$\Pi$ 是大写希腊字母 pi，这里表示现在）。此式与单一值公式的差别仅在于以 $R_t$ 代替了 $V$，并加入了 $\sum$ 符号。

和的思想很容易引入到连续现金流的情况，但在后一种情况下，$\sum$ 符号必定为定积分符号所代替。考察收益率为 $R(t)$ 美元的连续收入流。这意味着在 $t=t_1$，收益率为每年 $R(t_1)$ 美元，但在另一时点 $t=t_2$，收益率为每年 $R(t_2)$ 美元——视 $t$ 为连续变量。如果在任意时点 $t$，我们允许有无穷小时间 $dt$ 变化，在时期 $[t, t+dt]$ 的收益量可以写成 $R(t)dt$[参见前面对 $dK \equiv I(t)dt$ 的讨论]。当按年比率 $r$ 连续贴现时，其现值应为 $R(t)e^{-rt}dt$。若我们要求的是三年收入流的总现值，那么，可以通过如下定积分获得答案：

$$\Pi = \int_0^3 R(t)e^{-rt}dt \tag{13.11'}$$

463 此式是和(13.11)的连续形式，它与单一值公式的差别仅在于以 $R(t)$ 替换了 $V$，并加入了定积分符号。[①]

**例 6** 连续收入流按每年按 $D$ 美元不变收益率持续 $y$ 年，将其按年利息率 $r$ 贴现，其现值为多少？按照(13.11′)，有

$$\Pi = \int_0^y De^{-rt}dt = D\int_0^y e^{-rt}dt = D\left[\frac{-1}{r}e^{-rt}\right]_0^y$$

$$= \frac{-D}{r}e^{-rt}\Big]_{t=0}^{t=y} = \frac{-D}{r}(e^{-ry}-1) = \frac{D}{r}(1-e^{-ry}) \tag{13.12}$$

因此，$\Pi$ 取决于 $D$、$r$ 和 $y$。若 $D=3000$ 美元，$r=0.06$，$y=2$，我们有

$$\Pi = \frac{3000}{0.06}(1-e^{-0.12}) = 50,000(1-0.8869) = \$5655$$

[近似地]

$\Pi$ 值自然总为正；这是由 $D$、$r$ 及 $(1-e^{-ry})$ 为正推出来的。(数 $e$ 自乘至任何负幂总会产生一个正的分数值，由图 10.3$a$ 中的第 Ⅱ 象限可以看到这一点。)

**例 7** 在 10.6 节酒的窖藏问题中，我们假定窖减成本为零。当时采用这样一个简化假设，是因为我们不知道计算成本流量现值的方法。现在，我们已消除了这种无知，所以可以允许酒商发生窖藏成本。

令酒商现在发生的采购成本为 $C$；其未来销售额随时间变化而变化，一般可以表示成 $V(t)$，其现值为 $V(t)e^{-rt}$。虽然销售

---

① 应注意，和上限指标与积分上限指标在 3 是相同的，但和下限指标 1 与积分下限 0 不同。这是因为根据假设，离散收入流的第一笔收益直至 $t=1$(第一年末)方能收到，但连续情况下的收入流在 $t=0$ 以后便可以出现。

额仅是一个未来值(在本例中仅有一次销售交易),但窖藏成本则是一个支出流。假设成本是一个每年为固定比率 $s$ 美元的不变支出流,在 $t$ 年中所发生的窖藏成本的总现值等于

$$\int_0^t se^{-rt}dt = \frac{s}{r}(1-e^{-rt}) \qquad [参见(13.12)]$$

因此,酒商力求最大化的净现值可以表示成

$$N(t) = V(t)e^{-rt} - \frac{s}{r}(1-e^{-rt}) - C$$

$$= \left[V(t) + \frac{s}{r}\right]e^{-rt} - \frac{s}{r} - C$$

它是只有一个选择变量 $t$ 的目标函数。

为使 $N(t)$ 最大化,必须选择 $t$ 值以使 $N'(t)=0$。这个一阶 464 导数为

$$N'(t) = V'(t)e^{-rt} - r\left[V(t) + \frac{s}{r}\right]e^{-rt} [积的微分法则]$$

$$= [V'(t) - rV(t) - s]e^{-rt}$$

当且仅当

$$V'(t) = rV(t) + s$$

此一阶导数为零。因此,最后一个方程可以视为选择销售时间 $\bar{t}$ 的必要的最优化条件。

此条件的经济解释很容易直观推导出来:$V'(t)$ 表示销售额的变化率,或者 $V$ 的增量,若销售延迟一年,方程右边的两项分别表示由于这个延迟销售而导致的利息成本增量和窖藏成本增量(收益和成本均在时间 $t$ 计算)。所以,令方程两边相等的思想对我们而言不过是“新瓶装旧酒”,因为它只不过是条件 MC = MR 的另一种形式而已。

## 持久流量的现值

如果资金流量永远持续——比如从持久债券获得的利息或从如土地等恒久资产获得的收益——资金流的现值将为

$$\Pi = \int_0^{\infty} R(t) e^{-rt} dt$$

它是一个广义积分。

**例 8** 求每年按不变比率 $D$ 美元获得的恒久收入，按贴现率 $r$ 连续贴现的现值。因为在计算广义积分时，我们仅取正常积分的极限，所以(13.12)中的结果仍是有益的。具体地，我们可以写出

$$\Pi = \int_0^{\infty} De^{-rt} dt = \lim_{y \to \infty} \int_0^{y} De^{-rt} dt$$

$$= \lim_{y \to \infty} \frac{D}{r}(1 - e^{-ry}) = \frac{D}{r}$$

注意，表示年份的 $y$ 参数在最终答案中已消失。它也理应如此，因为我们这里讨论的是持久流量。读者也许可以注意到，我们的结果(现值 = 收益率/贴现率)恰好我们熟悉的持久收入领域中的所谓资产的“资本化”公式相对应。

## 练习 13.5

---

1 给定如下边际收益函数：

$$(a)\ R'(Q) = 28Q - e^{0.3Q} \quad (b)\ R'(Q) = 10(1 + Q)^{-2}$$

求每一题的总收益函数 $R(Q)$。为确定积分常数需引入什么初始条件？

465　2 $(a)$ 给定边际进口倾向 $M'(Y) = 0.1$，及当 $Y = 0$ 时 $M = 20$，求进

口函数 $M(Y)$。

(*b*) 给定边际消费倾向 $C'(Y)=0.8+0.1Y^{-1/2}$,以及当 $Y=100$ 时,$C=Y$,求消费函数 $C(Y)$。

3 假设以 $I(t)=12t^{1/3}$ 描述投资率,且 $K(0)=25$:

(*a*) 求资本存量的时间路径。

(*b*) 分别求时期[0,1]和[1,3]的资本积累量。

4 给定年固定收益率为 1000 美元的连续收入流:

(*a*) 若收入流持续 2 年,按年利率 0.05 连续贴现,那么,现值 $\Pi$ 为多少?

(*b*) 若收入流恰好在 3 年后终止,且贴现率为 0.04,现值 $\Pi$ 又为多少?

5 下列持久资金流量的现值为多少?

(*a*) 每年 1450 美元,按 $r=5\%$ 贴现。

(*b*) 每年 2460 美元,按 $r=8\%$ 贴现。

---

## 13.6 多马增长模型

在(13.1)和(13.2)的人口增长问题及在(13.10)中的资本形成问题中,共同的目标是在已知变量变化模式的基础上描述时间路径。而多马教授的古典增长模型[①] 的思想则是规定要满足某些均衡经济条件所需要的时间路径的类型。

### 多马增长模型的框架

多马模型的基本假设前提如下:

1. 年投资(流量)比率 $I(t)$ 的任意变化会产生双重效果:它

---

① E.D.多马:"资本扩张、增长率及就业",《经济计量学》,1946 年 4 月,137—147 页。重印于多马:《经济增长理论论文集》,牛津大学出版社,1957 年,70—82 页。

将影响总需求及该经济体的生产能力。

2. $I(t)$变化的需求效应通过乘数过程立即发挥作用。因此，$I(t)$的提高会通过$I(t)$增量的乘数作用增加年收入流量比率$Y(t)$。乘数是$k = 1/s$，其中$s$表示已知的边际储蓄倾向，假设$I(t)$是唯一的影响收入流量比率的（参数的）支出流量，则我们可以写出

466

$$\frac{dY}{dt} = \frac{dI}{dt}\frac{1}{s} \tag{13.13}$$

3. 投资的生产能力效应通过该经济能够生产的潜在产出能力的变化来度量。假定能力—资本比率不变，我们可以写出：

$$\frac{\kappa}{K} \equiv \rho \ (= a \text{ 常数})$$

其中$\kappa$（希腊字母 kappa）表示生产能力或每年的潜在产出流量，$\rho$（希腊字母 rho）表示已知资本—能力比率。当然，这意味着资本存量$K(t)$的经济体每年能够生产的产出或收入，等于$\kappa \equiv \rho K$美元。注意，由$d\kappa = \rho dK$（生产函数），可得$d_\kappa = \rho dK$，和

$$\frac{d\kappa}{dt} = \rho\frac{dK}{dt} = \rho I \tag{13.14}$$

在多马模型中，均衡被定义为生产能力得到充分利用的状态。因此，为达到均衡，要求总需求恰好等于该年度能够生产的潜在产出，即$Y = \kappa$。但是，若我们从均衡状态出发，则要求生产能力变化与总需求变化相等，即

$$\frac{dY}{dt} = \frac{d\kappa}{dt} \tag{13.15}$$

何种投资 $I(t)$的时间路径能够时时满足这一均衡条件？

**求　解**

为回答此问题，我们先将(13.13)和(13.14)代入均衡条件(13.15)。结果得到如下微分方程：

$$\frac{dI}{dt}\frac{1}{s}=\rho I \quad 或 \quad \frac{1}{I}\frac{dI}{dt}=\rho s \tag{13.16}$$

因为(13.16)设定了 $I$ 变化的确定模式，我们应当能够由此式求出均衡的(或所求的)投资路径。

在这个简单的例子中，直接将(13.16)中的第 II 个方程的两边对 $t$ 积分，便可得到解。事实上，方程两边相等，可以保证其积分也相等。因此，

$$\int \frac{1}{I}\frac{dI}{dt}dt=\int \rho s dt$$

根据代换法则和对数法则，方程左边给出 467

$$\int \frac{dI}{I}=\ln|I|+c_1(I\neq 0)$$

而右边则给出($\rho s$ 为常数)

$$\int \rho s dt=\rho st+c_2$$

令两个结果相等并合并常数，有

$$\ln|I|=\rho st+c \tag{13.17}$$

为从 $\ln|I|$求得$|I|$，我们要进行被称为“取反对数”的运算，这要利用公式 $e^{\ln x}=x$。令(13.17)的每一边成为常数 $e$ 的指数，得到

$$e^{\ln|I|}=e^{(\rho st+c)}$$

或 $$|I|=e^{\rho st}e^{c}=Ae^{\rho st} \quad 在此\ A\equiv e^{c}$$

若我们取投资为正，则$|I| = I$，从而上述结果成为 $I(t) = Ae^{\rho st}$，其中 $A$ 为任意值。为消去这个任意常数，令方程 $I(t) = Ae^{\rho st}$ 中的 $t = 0$，以得到 $I(0) = Ae^0 = A$。此式定义了常数 $A$，使我们可以将解，即所求的投资路径表示成

$$I(t) = I(0)e^{\rho st} \tag{13.18}$$

其中 $I(0)$表示初始投资率。[1]

此结果含有不平常的经济意义：为使生产能力和需求在不同时间保持平衡，投资流量比率必须按照指数 $\rho s$，沿着图 13.6 描述的路径增长。显然，所要求的投资增长率越高，生产能力—资本比率和边际储蓄倾向也应越大。但是，无论如何，一旦 $\rho$ 和 $s$ 的值已知，所要求的投资增长路径便完全确定了。

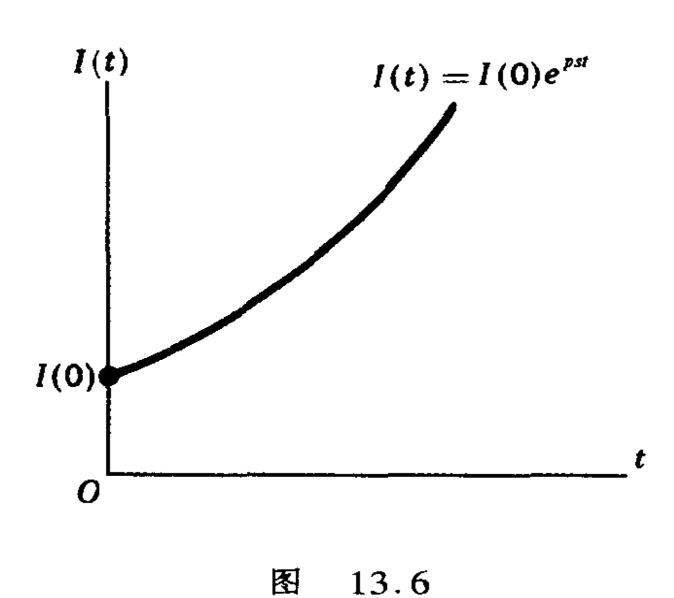



图 13.6

## 刃 锋

现在我们可以问这样一个问题：如果实际投资增长率（称其为

① 即使令结果$|I| = Ae^{\rho st}$中的投资为负，解(13.18)依然正确。参见练习 13.6－3。

比率 $r$)与所要求的比率 $\rho s$ 不同,会出现何种情况呢?

多马的方法是定义一个利用系数:

$$u = \lim_{t \to \infty} \frac{Y(t)}{\kappa(t)} \quad [u = 1 \text{ 意味着生产能力的充分利用}]$$

并证明 $u = r/\rho s$,从而当 $r \gtrless \rho s$ 时,$u \gtrless 1$。换言之,如果实际的和所要求的比率间存在差距($r \neq \rho s$),那么,我们最终将发现(当 $t \to \infty$)或者生产能力不足($u > 1$),或者生产能力过剩($u < 1$),具体情况视 $r$ 是大于还是小于 $\rho s$ 而定。

但是,我们可以证明,关于生产能力短缺或过剩的结论可以在任意时间 $t$ 应用,而不是仅在 $t \to \infty$ 时才能应用。增长率为 $r$ 意味着

$$I(t) = I(0)e^{rt} \quad \text{和} \quad \frac{dI}{dt} = rI(0)e^{rt}$$

因此,由(13.13)和(13.14),我们有

$$\frac{dY}{dt} = \frac{1}{s}\frac{dI}{dt} = \frac{r}{s}I(0)e^{rt}$$

$$\frac{d\kappa}{dt} = \rho I(t) = \rho I(0)e^{rt}$$

这两个导数间的比率

$$\frac{dY/dt}{d\kappa/dt} = \frac{r}{\rho s}$$

应当能够揭示出在实际增长率为 $r$ 条件下,在任意时间 $t$,投资的需求创造效应与投资的生产能力生成效应的相对大小。如果实际投资增长率 $r$ 大于 $\rho s$,则 $dY/dt > d\kappa/dt$,那么,需求效应将超过生产能力效应,导致生产能力不足。相反,若 $r < \rho s$,将存在总需求不足,从而导致生产能力过剩。

这一结论的新奇之处在于：如果投资的实际增长率快于所要求的比率($r>\rho s$)，那么，最终将导致生产能力短缺而非过剩。同样令人费解的是：如果实际投资增长滞后于所要求的比率($r<\rho s$)，我们将面临生产能力过剩而非短缺。的确，由于这样一个相悖的结果，如果我们现在让厂商按照其目前的生产能力状况调整其实际投资增长率 $r$（到目前为止，视 $r$ 为常量），厂商多半会做出469“错误”的调整决策。比如，在 $r>\rho s$ 情况下，严重的生产能力短缺会刺激更快的投资增长率，而这意味着 $r$ 的增加，而非在此情形下所要求的 $r$ 的减少。结果，两个增长率之间的偏差不仅不会缩小，而且还会扩大。

最终的结果是：给定参常数 $\rho$ 和 $s$，则避免生产能力短缺或过剩的唯一方法是谨慎地控制投资流，使其沿着均衡路径，以增长率 $\bar{r}=\rho s$ 增长。而且，正如我们已经证明的那样，任何对这个“刃锋”时间路径的偏离都将导致多马在此模型中所设想的生产能力充分利用标准永远不能得到满足。这个前景可能不太令人感到鼓舞。幸运的是，如果对多马模型中的某些假定进行修正，则可能得到一些更为灵活的结论。我们会从下一章将要讨论的索罗教授的增长模型中看到这一点。

## 练习 13.6

---

1 多马模型明确考察了多少种生产要素？对生产中的资本劳动比率而言，这一事实意味着什么？

2 在 10.2 节中我们知道，指数函数 $Ae^{rt}$ 中的常数 $r$ 表示函数的增长

率。将其应用于(13.16)，并在不通过积分的情况下导出(13.18)。

3 证明：即使令方程 $|I| = Ae^{\rho st}$ 中的投资为负，但通过定义任意常数 $A$，我们仍可得到最后的解(13.18)。

4 证明：(13.18)中的结论还可以通过另一种方式得到，即以 $t=0$ 和 $t=t$ 为积分限，将(13.16)的两边

$$\frac{1}{I}\frac{dI}{dt} = \rho s$$

对变量 $t$ 求定积分，并令两边相等。记住，当我们将积分变量由 $t$ 变为 $I$ 时，积分限将由 $t=0$ 和 $t=t$ 分别变为 $I=I(0)$ 和 $I=I(t)$。

# 第 14 章 连续时间：一阶微分方程

470

在多马增长模型中，我们已通过直接积分解了一个简单的微分方程。对于更复杂的微分，也有各种现成的解法。但是，即便在后一种情况下，微分方程解法依赖的基本思想依然是积分的方法。因此，解微分方程通常也称为该方程的**积分**。

本章仅讨论**一阶**微分方程。在这里，“**阶**”是指在微分方程中所出现的导数（或微分）的最高阶数。因此，一阶微分方程仅包含一阶导数，比如 $dy/dt$。

## 14.1 具有常系数和常数项的一阶线性微分方程

一阶导数 $dy/dt$ 是在一阶微分方程中出现的唯一导数，但它可能以不同的幂数出现：$dy/dt$，$(dy/dt)^2$ 或 $(dy/dt)^3$。方程中导数所达到的最高幂数称为微分方程的**次**。在导数 $dy/dt$ 仅为一次，因变量 $y$ 也是一次，而且没有积 $y\cdot(dy/dt)$ 等形式出现的情况下，此方程便称为**线性的**。因此，一阶线性微分方程的一般形

471

式为[1]

$$\frac{dy}{dt} + u(t)y = w(t) \tag{14.1}$$

其中 $u$ 和 $w$ 同 $y$ 一样，都是 $t$ 的函数。但是，与 $dy/dt$ 和 $y$ 相反，对自变量 $t$ 没有任何限制。因此，函数 $u$ 和 $w$ 可以非常好地表示诸如 $t^2$、$e^t$ 或 $t$ 的更复杂的函数。另一方面，$u$ 和 $w$ 也可以表示常数。

上述内容的最后一点引导我们对方程作进一步分类。当函数 $u$（因变量 $y$ 的系数）为常数，且当函数 $w$ 是一个相加性常数项时，(14.1)简化为具有常数系数和常数项的一阶线性微分方程的特殊情况。本节，我们将只讨论这类简单的微分方程。

## 齐次方程的情况

若 $u$ 和 $w$ 为常函数，且如果 $w$ 恰好恒为零，(14.1)将变成

$$\frac{dy}{dt} + ay = 0 \tag{14.2}$$

其中 $a$ 为一常数。考虑到常数项为零，此微分方程被称作齐次方程（与齐次方程组相比较）。更确切地讲，此方程是齐次方程的原因在于其中的 $y$ 和 $dy/dt$ 项均是一次的。特别地，有别于其它项的常数项零，可以视为 $y$ 的一次幂，因为 $0y=0$。

方程(14.2)还可以写成

$$\frac{1}{y}\frac{dy}{dt} = -a \tag{14.2'}$$

---

[1] 注意，(14.1)中的导数项 $dy/dt$ 具有单位系数。这并不意味着永远不会有不是 1 的系数，然而，当不为 1 的系数出现时，我们总可以用所说的系数相除而使方程“正规化”。因此，(14.1)给出的方程可以视为一个一般的表达式。

但读者应认识到，我们在多马模型中遇到的微分方程(13.16)恰好也属于这种形式。因此，类似地，我们可以立即将(14.2)和(14.2′)的解写出如下：

$$y(t)=Ae^{-at} \qquad [\text{通解}] \tag{14.3}$$

或

$$y(t)=y(0)e^{-at} \qquad [\text{特解}] \tag{14.3'}$$

472 在(14.3)中出现一个任意常数 $A$；因此，它是一个**通解**。当 $A$ 为任意具体值代替时，此通解便成为(14.2)的特解。此特解有无数个，对于每个可能的 $A$ 值，包括 $y(0)$，都有一个特解。但是最后一个值具有特殊意义：$y(0)$是使解满足初始条件的唯一值。因为它表示使任意常数确定化的结果，所以，我们将(14.3′)称作微分方程(14.2)或(14.2′)的**定解**。

关于微分方程的解，读者应观测到两点：(1)解不是一个数值，而是一个函数 $y(t)$——若 $t$ 表示时间，那么，它表示时间路径。(2)解 $y(t)$不含有任何导数或微分表达式，所以只要将 $t$ 的具体值代入此解，就可以直接算出相应的 $y$ 值。

### 非齐次函数的情况

当(14.2)中的零为非零常数所取代，我们便得到**非齐次**线性微分方程：

$$\frac{dy}{dt}+ay=b \tag{14.4}$$

此方程的解将由两项之和构成，其中一项称作**余函数**，以 $y_c$ 表示；另一项称作**特别积分**，以 $y_p$ 表示。后面将要表明，这两项均具有重要的经济意义。这里我们将仅给出方程的解法，关于解法的合

理性，后面自然会清楚。

尽管我们的目标是解非齐次方程(14.4)，但我们要常常将其视为如(14.2)所示的齐次形式。为便于表述，我们将其称作(14.4)的**简化**方程。相应地，(14.4)本身被称作**完备方程**，由此可知，余函数 $y_c$ 只不过是简化方程的通解，而特别积分 $y_p$ 只是完备方程的**任意**特解。

前面对齐次方程情况已经给出了简化方程的通解，因而我们可以给出

$$y_c = Ae^{-at} [\text{由}(14.3)]$$

那么，特别积分又如何呢？因为特别积分是完备方程的任意特解，所以我们首先试求最简单的可能的解，即 $y$ 为某常数($y=k$)时的解。若 $y$ 是常数，则由此得 $dy/dt=0$，且(14.4)成为 $ay=b$，有解 $y=b/a$。因此，只要 $a\neq 0$，常数解便成立。在此情况下，我们有

$$y_p = \frac{b}{a} \quad (a \neq 0)$$

则余函数和特别积分的和构成了完备函数(14.4)的通解： 473

$$y(t) = y_c + y_p = Ae^{-at} + \frac{b}{a} \quad [\text{通解}, a \neq 0 \text{ 的情况}] \tag{14.5}$$

上式成为通解是由于任意常数 $A$ 的存在。当然，我们可以利用初始条件确定这个常数。当 $t=0$ 时，令 $y$ 的值为 $y(0)$。则令(14.5)中的 $t=0$，可求得

$$y(0) = A + \frac{b}{a} \text{ 和 } A = y(0) - \frac{b}{a}$$

因此，可将(14.5)重写成

$$y(t) = \left[y(0) - \frac{b}{a}\right]e^{-at} + \frac{b}{a}$$

[特解，$a \neq 0$ 的情况] (14.5′)

应注意，利用初始条件确定任意常数是——而且应当是在求得完备方程的通解后的最后一个步骤。因为 $y_c$ 和 $y_p$ 的值均与 $y(0)$ 的值相联系，所以，在确定常数 $A$ 时，必须将二者都考虑进去。

**例 1** 解方程 $dy/dt+2y=6$，初始条件 $y(0)=10$。这里，我们有 $a=2, b=6$，因此，由(14.5′)，解为

$$y(t) = [10 - 3]e^{-2t} + 3 = 7e^{-2t} + 3$$

**例 2** 解方程 $dy/dt+4y=0$，初始条件 $y(0)=1$。因为 $a=4, b=0$，我们有

$$y(t) = [1 - 0]e^{-4t} + 0 = e^{-4t}$$

由(14.3′)可以得到同样的答案；(14.3′)属于齐次函数的情况。齐次方程(14.2)只是非齐次方程(14.4)当 $b=0$ 时的一个特例。因而，公式(14.3′)也是公式(14.5′)在 $b=0$ 情况下的一个特例。

如果 $a=0$，从而解(14.5′)未有定义，会怎么样呢？在此情况下，微分方程是一种极简单的形式。

$$\frac{dy}{dt} = b \tag{14.6}$$

通过直接积分，可以求得其通解为

$$y(t) = bt + c \tag{14.7}$$

其中 $c$ 为任意常数。事实上，(14.7)中的两项也可以分别被视为已知微分方程的余函数和特别积分。因为 $a=0$，余函数可以表示成

$$y_c = Ae^{-at} = Ae^0 = A \ (A = 任意常数)$$

至于特别积分,在本例 $a=0$ 的情况下,常数解 $y=k$ 不成立这一 474
事实表明,我们将会求得一个非常数解。下面,我们考察一种最简单的类型,即 $y=kt$ 的情况。若 $y=kt$,则 $dy/dt=k$,完备方程(14.6)将简化为 $k=b$,从而可以写出

$$y_p = bt \quad (a=0)$$

我们的新的试探解的确成立!因而,(14.6)的通解为

$$y(t) = y_c + y_p = A + bt \quad [\text{通解}, a=0 \text{ 的情况}] \tag{14.7'}$$

它与(14.7)的结果是一致的,因为 $c$ 和 $A$ 只不过是任意常数的另一个符号而已。但要注意,在本例中,$y_c$ 是一个常数,而 $y_p$ 是时间的函数,这恰与(14.5)中的情形相反。

通过确定任意常数,可以求得定解为

$$y(t) = y(0) + bt \quad [\text{特解}, a=0 \text{ 的情况}] \tag{14.7''}$$

**例3** 解方程 $dy/dt=2$,初始条件 $y(0)=5$。由(14.7″),解为

$$y(t) = 5 + 2t$$

## 解的检验

的确,所有微分方程的解的正确性都可以通过微分来检验。

如果要对(14.5′)进行检验,我们可以得到导数

$$\frac{dy}{dt} = -a\left[y(0) - \frac{b}{a}\right]e^{-at}$$

如果将此 $dy/dt$ 表达式和(15.5′)中所示的 $y(t)$ 表达式代入微分方程(14.4)的左边,如果此方程的解正确,则代入上述表达式的左边应当恰好等于(14.4)右边的常数项 $b$。进行上述替代后,我们的确得到

$$-a\left[y(0)-\frac{b}{a}\right]e^{-at}+a\left\{\left[y(0)-\frac{b}{a}\right]e^{-at}+\frac{b}{a}\right\}=b$$

因此，如果此解也满足初始条件，那么，它便是正确的。为对它是否满足初始条件，我们令(14.5′)中的 $t=0$。因为

$$y(0)=\left[y(0)-\frac{b}{a}\right]+\frac{b}{a}=y(0)$$

是一个恒等式，所以确定满足初始条件。

应注意的是，作为解微分方程的最后一个步骤，读者应通过如下习惯检验答案的正确性：(1)保证时间路径 $y(t)$ 的导数与已知微分方程相一致。(2)确保定解满足初始条件。

## 475 练习 14.1

---

1 给定如下微分方程，求 $y_c$，$y_p$，通解，定解：

(a) $\frac{dy}{dt}+4y=12$; $y(0)=2$

(b) $\frac{dy}{dt}-2y=0$; $y(0)=9$

(c) $\frac{dy}{dt}+10y=15$; $y(0)=0$

(d) $2\frac{dy}{dt}+4y=6$; $y(0)=1\frac{1}{2}$

2 检验上题答案的正确性。

3 运用之中提出的适当公式，求下列每个方程的解：

(a) $\frac{dy}{dt}+y=4$; $y(0)=0$

(b) $\frac{dy}{dt}=23$; $y(0)=1$

(c) $\frac{dy}{dt}-5y=0$; $y(0)=6$

(d) $\frac{dy}{dt}+3y=2$; $y(0)=4$

(e) $\frac{dy}{dt}-7y=7$; $y(0)=7$

(f) $3\frac{dy}{dt}+6y=5$; $y(0)=0$

4 检验上题答案的正确性。

---

## 14.2 市场价格的动态学

在(宏观)多马增长模型中,我们认识了一阶线性齐次微分方程的一个应用。为介绍非齐次方程的应用,我们给出一个(微观的)动态市场模型。

### 框 架

对某一特定商品,假设其需求与供给函数如下

$$Q_d = \alpha - \beta P \qquad (\alpha, \beta > 0)$$

$$Q_s = -\gamma + \delta P \qquad (\gamma, \delta > 0) \tag{14.8}$$

则根据(3.4),均衡价格应为[①]

$$\bar{P} = \frac{\alpha + \gamma}{\beta + \delta} \text{(某一确定常数)} \tag{14.9}$$

如果初始价格 $P(0)$ 恰好在 $\bar{P}$ 水平,市场显然处于早已达到的均衡状态,勿需进行动态分析。但是,在 $P(0) \neq \bar{P}$ 的更重要的情况下,$\bar{P}$ 仅在经过适当的调整过程之后才能达到;在调整过程中,不仅价格随时间变化,而且,作为价格 $P$ 的函数,$Q_d$ 和 $Q_s$ 也必然随时间的变化而变化。从这个角度看,价格和数量变量都可以视为时间的函数。

我们的动态问题是:给定调整过程所需要的充分时间,能够将 476 价格调整至均衡水平 $\bar{P}$ 吗?亦即,当 $t \to \infty$ 时,时间路径 $P(t)$ 趋

① 我们已将(3.4)中的符号($a$、$b$、$c$、$d$)变换成($\alpha$、$\beta$、$\gamma$、$\delta$),以避免与使用 $a$ 和 $b$ 为参数的微分方程相混淆。我们在这里将把(14.4)应用于市场模型。

向收敛于 $\bar{P}$ 吗?

**时间路径**

为回答上述问题,我们必须先求出时间路径 $P(t)$。而这又要求我们先描述价格变化的具体形式。一般而言,价格变化是由市场中供给和需求的相对强度决定的。为简化起见。我们假设在某一时刻价格对时间的变化率总是在该时刻存在的超额需求的比例。这个价格变化模式可用公式表示成

$$\frac{dP}{dt} = j(Q_d - Q_s)\ (j > 0) \tag{14.10}$$

其中 $j$ 表示不变调整常数。根据这个变化模式,当且仅当 $Q_d = Q_s$ 时,我们有 $dP/dt = 0$。在这方面,注意到术语均衡价格的两种含义是有启发的:瞬时意义和市场出清意义(均衡价格是使 $Q_d$ 和 $Q_s$ 相等的价格)。在现在的模型中,这两种含义恰好是相互重合的,但在其它模型却未必如此。

根据(14.8)的需求和供给函数,我们可以将(14.10)具体表示成如下形式:

$$\frac{dP}{dt} = j(\alpha - \beta P + \gamma - \delta P) = j(\alpha + \gamma) - j(\beta + \delta)P$$

或

$$\frac{dP}{dt} + j(\beta + \delta)P = j(\alpha + \gamma) \tag{14.10'}$$

因为它恰好是微分方程(14.4)的形式,且因为 $P$ 的系数为非零,我们可以应用解的公式(14.5′),并将解——价格的时间路径——写成

$$P(t)=[P(0)-\frac{\alpha+\gamma}{\beta+\delta}]e^{-j(\beta+\delta)t}+\frac{\alpha+\gamma}{\beta+\delta}$$

$$=[P(0)-\bar{P}]e^{-kt}+\bar{P} \quad [\text{由}(14.9);k\equiv j(\beta+\delta)] \tag{14.11}$$

## 均衡的动态稳定性

这样，最初提出的问题，即当 $t\to\infty$ 时，$P(t)$是否趋近于 $\bar{P}$，相当于这样一个问题：当 $t\to\infty$ 时，(14.11)右边的第一项是否趋近于零？因为 $P(0)$和 $\bar{P}$ 均为常数，所以关键因素是指数表达式 $e^{-kt}$。考虑到 $k>0$，当 $t\to\infty$ 时，此指数式确定趋于零。因而，在 477 模型假设的基础上，时间路径确实使价格趋于均衡状态。在相关变量 $P(t)$的时间路径收敛于 $\bar{P}$ 水平（这里将其解释成瞬时均衡而非市场出清均衡）这类情形中，我们把这种均衡称为**动态稳定**的均衡。

动态稳定性是一个非常重要的概念。下面，我们对(14.11)作更详尽的分析，以便对这一概念作进一步考察。解(14.11)依赖于 $P(0)$和 $\bar{P}$的相对大小，包含三种可能的情况。第一种情况 $P(0)=\bar{P}$，这意味着 $P(t)=\bar{P}$。在此情况下，价格的时间路径可像图 14.1 那样，绘成一条水平的直线。如前面所提到的那样，在这种情况下均衡可立即达到。第二种情况是 $P(0)>\bar{P}$。在此情况下，(14.11)右边第一项为正，但当 $t$ 增加使 $e^{-kt}$值下降时，第一项的值将减小。因此，时间路径将从上方趋向于均衡水平 $\bar{P}$，其情形如 14.1 最上边一条曲线所示。第三种情况与第二种情况恰好相反，$P(0)<\bar{P}$；如图 14.1 最下面一条曲线所示，时间路径将从下方趋向于均衡价格水平 $\bar{P}$。一般而言，要具备动态稳定性，

时间路径与均衡的偏差，或者是等于零（如第一种情况），或者随时间而递减（如第二、三种情况）。

(14.11)与(14.5′)的比较表明，对应于 $b/a$ 的 $\bar{P}$，项，只不过是特别积分 $y_p$，而指数是（已确定的）余函数 $y_c$。因此，我们现在有了 $y_c$ 和 $y_p$ 的经济解释：$y_p$ 表示相关变量的瞬时均衡水平，$y_c$ 表示均衡偏差。动态稳定性要求当 $t\to\infty$ 时，余函数渐近为零。

在此模型中，特别积分为常数，从而有瞬时意义上的稳定均衡，以 $\bar{P}$ 表示。如果特别积分不是常数，如(14.7′)中那样，那么，我们可以将其解释为移动均衡。

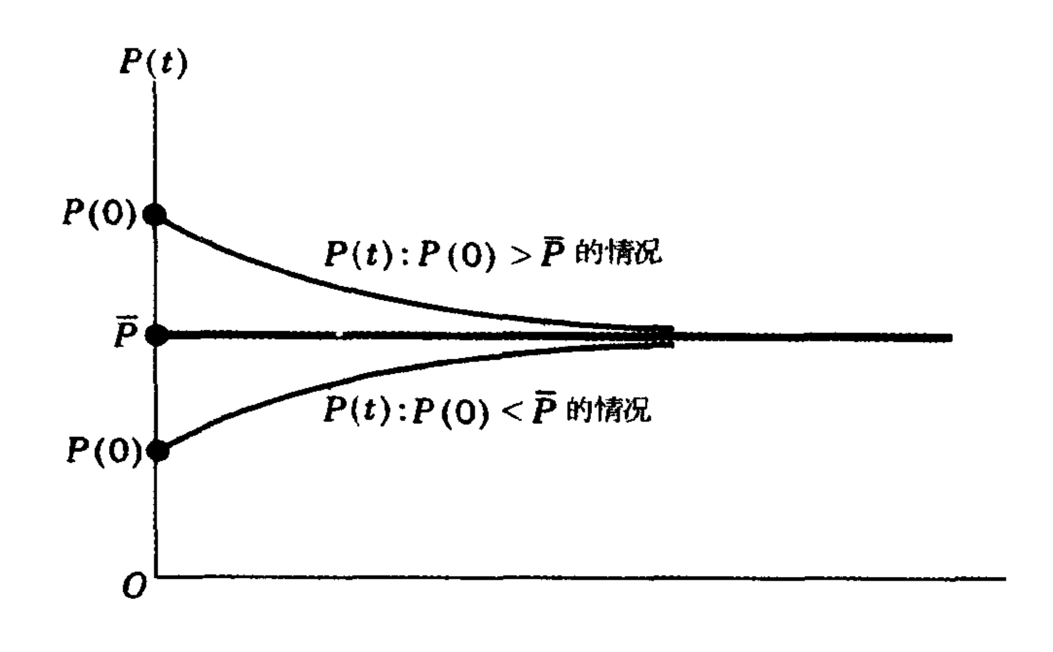



图 14.1

478

## 模型的另一种应用

我们上面所做的工作是在给定某些参数符号的情况下，分析均衡的动态稳定性（时间路径的收敛性）。另一类问题是：要保证动态稳定性，对参数要施加何种具体限制？

此问题的答案包含在解(14.11)中。如果我们允许 $P(0)\neq$

$\bar{P}$,可以看到当且仅当 $k>0$,即当且仅当

$$j(\beta+\delta)>0$$

时,在 $t\to\infty$ 时,(14.11)中的第一项($y_c$)才趋于零。因此,我们可以将上述不等式作为对参数 j(价格调整系数),β(需求曲线斜率的)负值,其中 $Q$ 是纵轴),δ(供给曲线的斜率,同样把 $Q$ 绘成纵轴)的限制。

在价格调整为"正常"的情形,即当 $j>0$,超额需求将驱使价格上升而非下降的情况下,这个约束意味着(β+δ)>0,或者,等价地

$$\delta>-\beta$$

在此情况下,为实现动态稳定性,供给曲线的斜率必须超过需求曲线的斜率。当供给曲线和需求曲线如图(14.8)那样,是正常倾斜(−β<0,δ>0)时,这个要求显然是满足的。但是,即使其中的一条曲线倾斜方向"反常",如当 δ=1 和 −β=1/2(斜率为正的需求曲线)时,这个条件仍可能得到满足。图 14.2 描述了这种情况,其中均衡价格 $\bar{P}$同以往一样,仍由两条曲线的交点决定。如果初始价格恰好在 $P_1$,则 $Q_d$(距离 $P_1G$)将超过 $Q_s$(距离 $P_1F$),并且超额需求($FG$)将驱使价格上升。另一方面,若初始价格在 $P_2$,则存在负的超额需求 $MN$,这将驱使价格下降。因此,如图中的两个箭号所示,价格调整在此情况下将趋于均衡,而与从 $\bar{P}$的哪一边开始无关。但我们应当强调,尽管箭号可以表明调整方向,却不能表明变化的大小。因此,图 14.2 从本质上看基本上是静态的,而不是动态的,它只能用于说明而不是替代上面所提出的动态分析。 479

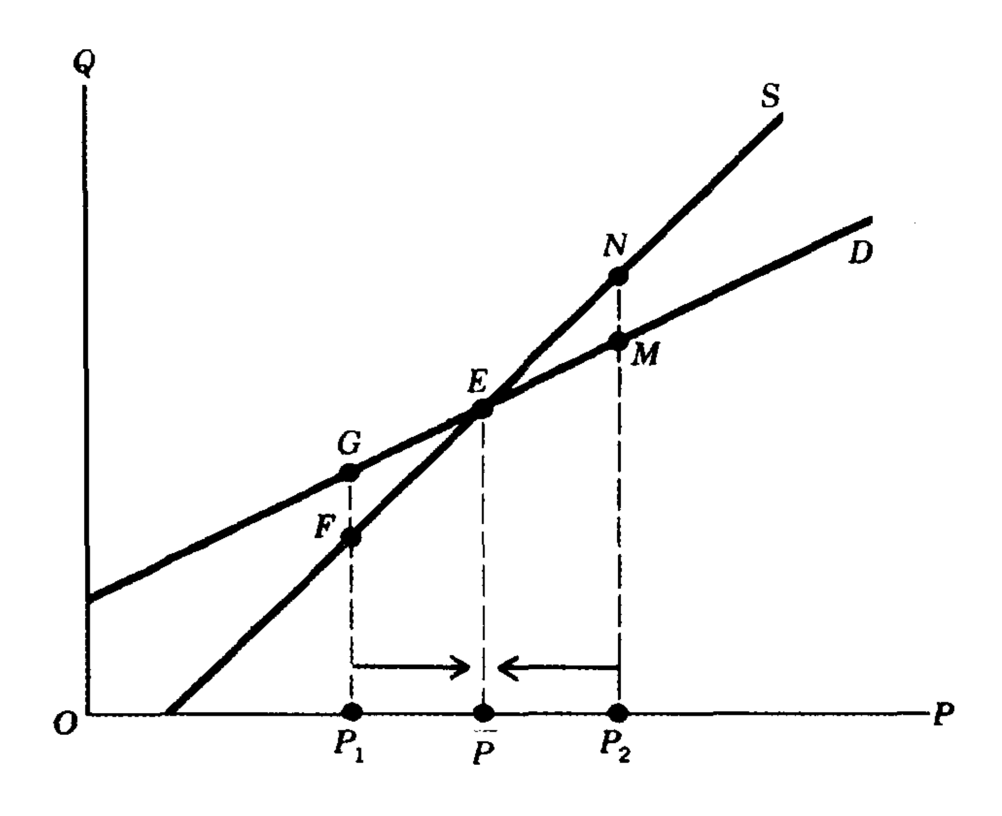



图 14.2

## 练习 14.2

---

1 如果图 14.2 中的需求曲线和供给曲线的斜率均为负,要具备动态稳定性,哪条曲线应更陡一些?你的答案是否与判别标准 $\delta > -\beta$ 一致?

2 证明(14.10′)可重写成 $dP/dt + k(P - \bar{P}) = 0$。如果令 $P - \bar{P} \equiv \Delta$(表示偏差),则有 $d\Delta/dt = dP/dt$,进而微分方程可以重写成

$$\frac{d\Delta}{dt} + k\Delta = 0$$

求时间路径 $\Delta(t)$,讨论动态稳定性的条件。

3 本节所讨论的动态市场模型是模仿 3.2 节的静态模型所构建的。将静态模型转化为动态模型需要何种新的特征?

4 令需求与供给函数为

$$Q_d = \alpha - \beta P + \sigma \frac{dP}{dt} \quad Q_s = -\gamma + \delta P \qquad (\alpha, \beta, \gamma, \delta > 0)$$

($a$) 假设价格对时间的变化率恰好是超额需求的比例,求时间路径 $P$

$(t)$(通解)。

(b) 何为瞬时均衡价格？何为市场出清均衡价格？

(c) 要保证动态稳定性，应对参数 $\sigma$ 施加各种限制？

5 令需求与供给函数为

$$Q_d = \alpha - \beta P - \eta \frac{dP}{dt} \qquad Q_s = \delta P \qquad (\alpha, \beta, \eta, \delta > 0)$$

(a) 假设市场在每一时点都是出清的，求时间路径 $P(t)$。

(b) 该市场是否具有动态稳定的瞬时均衡价格？

(c) 现在的模型中，对所有的 $t$，$Q_d = Q_s$ 的假设与 3.2 节中静态市场模型的假设是一致的。但这里的模型仍是动态模型。为什么？

---

## 14.3 可变系数和可变项

480

在更一般的一阶线性微分方程中

$$\frac{dy}{dt} + u(t)y = w(t) \tag{14.12}$$

$u(t)$和 $w(t)$分别表示可变系数和可变项。在这种情况下，如何求出时间路径 $y(t)$？

### 齐次方程的情况

对于齐次方程的情况，其中 $w(t)=0$，方程的解很容易求出。因为微分方程的形式为

$$\frac{dy}{dt} + u(t)y = 0 \qquad \text{或} \qquad \frac{1}{y}\frac{dy}{dt} = -u(t) \tag{14.13}$$

两边依次对 $t$ 积分，得到

$$\text{左边} = \int \frac{1}{y}\frac{dy}{dt}dt = \int \frac{dy}{y} = \ln y + c \text{ (假定 } y > 0)$$

$$右边 = \int -u(t)dt = -\int u(t)dt$$

后一个方程难以进行进一步的积分，因为 $u(t)$ 未给出具体形式。因此，我们不得不满足于一个一般积分表达式。当方程两边相等时，结果成为：

$$\ln y = -c - \int u(t)dt$$

则所求的 $y$ 路径可以通过求 $\ln y$ 的反对数得到：

$$y(t) = e^{\ln y} = e^{-c}e^{-\int u(t)dt} = Ae^{-\int u(t)dt} \quad 在此\ A \equiv e^{-c} \tag{14.14}$$

这是微分方程(14.13)的通解。

为突出系数 $u(t)$ 的可变性质，我们已明确地写出了变量 $t$。但为了简化书写符号，我们从现在起省略自变量，并将 $u(t)$ 简化为 $u$。

将不变系数模型的通解(14.3)与(14.14)相比较，(14.14)中唯一的修正是以更复杂的表达式 $e^{-\int u\,dt}$ 代替了 $e^{-at}$。如果我们将 $e^{-at}$ 中的 $at$ 解释成 $\int a\,dt = at$（加上一个可纳入 A 项中的常数，因为 $e$ 自乘常数幂仍为常数），我们可以更好地理解这种变化的合理性。由此看来，这种差别就变成相似性了。因为在两种情况下，我们都取微分方程中 $y$ 项的系数（在一种情况中为常数项 $a$，在另一种情况中为可变项 $u$）并将 $y$ 对 $t$ 积分，然后再取所得积分的负值作为 $e$ 的指数。

一旦得到通解，根据适当的初始条件求得定解，便是一个相对简单的事了。

481 **例 1** 求方程 $\frac{dy}{dt} + 3t^2y = 0$。这里 $u = 3t^2$，且 $\int udt = \int 3t^2dt$

$= t^2 + c$。因此，由(14.14)，我们可以把解写成

$$y(t) = Ae^{-(t^3+c)} = Ae^{-t^3}e^{-c} = Be^{-t^3} \qquad \text{在此 } B \equiv Ae^{-c}$$

注意，如果省去积分常数 $c$，我们不会失去任何信息，因为那样我们会得到 $y(t) = Ae^{-t^3}$，它与上式是相同的解，因为 $A$ 和 $B$ 均表示任意常数。换言之，指数式 $e^{-c}$（其中常数仅出现常数 $c$）总可以归入另一个常数 $A$ 中。

## 非齐次方程的情况

对于非齐次方程的情况，其中 $w(t) \neq 0$，解方程就有些难度了。我们将通过下一节将要讨论的恰当微分概念试求其解。但在这里先给出结果并无任何害处：给定微分方程(14.12)，通解为

$$y(t) = e^{-\int u\,dt}\left(A + \int we^{\int u\,dt}dt\right) \qquad (14.15)$$

其中 $A$ 为如果具有适当的初始条件，便可以确定的任意常数。

有趣的是，这个通解像不变系数、不变项一样，也包含两个相加的项。而且两项中的一项，$Ae^{-\int u\,dt}$，不过是前面导出的简化（齐次）方程的通解(14.14)，因此实质上是一个余函数。

**例 2** 求方程 $\frac{dy}{dt} + 2ty = t$ 的通解。因此我们有

$$u = 2t \quad w = t \quad \text{和} \quad \int u\,dt = t^2 + k \qquad (k \text{ 任意常数})$$

所以由(14.15)，有

$$y(t) = e^{-(t^2+k)}\left(A + \int te^{t^2+k}dt\right)$$

$$= e^{-t^2}e^{-k}\left(A + e^k\int te^{t^2}dt\right)$$

$$= Ae^{-k}e^{-t^2} + e^{-t^2}\left(\frac{1}{2}e^{t^2} + c\right) \quad [e^{-k}e^{k} = 1]$$

$$= (Ae^{-k} + c)e^{-t^2} + \frac{1}{2}$$

$$= Be^{-t^2} + \frac{1}{2} \qquad \text{在此 } B \equiv Ae^{-k} + c \text{ 是任意常数}$$

此解的正确性也可通过微分来检验。

482　注意到这一点是有趣的：在本例中，我们仍可省略积分常数 $k$ 及积分常数 $c$，而不影响最终结果，这是因为 $k$ 和 $c$ 均可归入最终解中的任意常数 $B$。希望读者试验不使用常数 $k$ 和 $c$ 的应用(14.15)的过程，并验证会得到相同的解。

**例 3**　解方程 $\frac{dy}{dt} + 4ty = 4t$。这次，我们将省去积分常数。因为

$$u = 4t \quad w = 4t \text{ 和} \int u\, dt = 2t^2 \quad [\text{省去常数}]$$

由(14.15)，通解为

$$y(t) = e^{-2t^2}\left(A + \int 4te^{2t^2}dt\right) = e^{-2t^2}(A + e^{2t^2}) \quad [\text{省去常数}]$$

$$= Ae^{-2t^2} + 1$$

正如所预期的那样，省略积分常数可以极大地简化运算步骤。

(14.12)中的微分方程 $\frac{dy}{dt} + uy = w$ 是一个比 $\frac{dy}{dt} + ay = b$ (14.4)更具一般性的微分方程，因为 $u$ 和 $w$ 不必像 $a$ 和 $b$ 那样，一定为常数。相应地，解的公式(14.15)也是比(14.5)更一般的解。事实上，当我们令 $u = a, w = b$ 时，应当能把(14.15)简化为(14.5)。事实也确实如此。因为那时，我们有

$$u = a \quad w = b \text{ 和} \int u\,dt = at \qquad [\text{省去常数}]$$

则(14.15)变成

$$y(t) = e^{-at}\left(A + \int be^{at}dt\right) = e^{-at}\left(A + \frac{b}{a}e^{at}\right) \qquad [\text{省去常数}]$$

$$= Ae^{-at} + \frac{b}{a}$$

它与(14.5)是一致的。

## 练习 14.3

---

解下列一阶线性微分方程。如果给定初始条件,确定任意常数:

1 $\frac{dy}{dt} + 5y = 15$

2 $\frac{dy}{dt} + 2ty = 0$

3 $\frac{dy}{dt} + 2ty = t$; $y(0) = \frac{3}{2}$

4 $\frac{dy}{dt} + t^2y = 5t^2$; $y(0) = 6$

5 $2\frac{dy}{dt} + 12y + 2e^t = 0$; $y(0) = \frac{6}{7}$

6 $\frac{dy}{dt} + y = t$

---

483

## 14.4 恰当微分方程

我们现在将引入恰当微分方程的概念,并运用这种解法解微分方程(14.12)以得到解的公式(14.15)。尽管我们现在的目的是

解线性微分方程，但恰当微分方程本身可以是线性的，也可以是非线性的。

## 恰当微分方程

给定二元函数 $F(y,t)$，其全微分为

$$dF(y,t) = \frac{\partial F}{\partial y}dy + \frac{\partial F}{\partial t}dt$$

令此微分等于零，所得到的方程

$$\frac{\partial F}{\partial y}dy + \frac{\partial F}{\partial t}dt = 0$$

被称作恰当微分方程，因为其左边恰好是 $F(y,t)$ 的微分。例如，给定

$$F(y,t) = y^2 t + k \qquad (k \text{ 常数})$$

全微分为

$$dF = 2yt\,dy + y^2 dt$$

因此微分方程

$$2yt\,dy + y^2 dt = 0 \qquad \text{或} \qquad \frac{dy}{dt} + \frac{y^2}{2yt} = 0 \tag{14.16}$$

是恰当微分方程。

一般而言，微分方程

$$M\,dy + N\,dt = 0 \tag{14.17}$$

当且仅当存在一个函数 $F(y,t)$ 使得 $M = \partial F/\partial y$ 和 $N = \partial F/\partial t$ 时，便是恰当的。然而，根据杨氏定理 $\partial^2 F/\partial t\,\partial y = \partial^2 F/\partial y\,\partial t$，

484 我们还可以表明，当且仅当

$$\frac{\partial M}{\partial t} = \frac{\partial N}{\partial y} \tag{14.18}$$

时，(14.17)是恰当的。上面的方程使我们得以对微分方程的恰当性进行简单的检验。将其应用于(14.16)，其中 $M=2yt, N=y^2$，此检验产生 $\partial M/\partial t=2y=\partial N/\partial y$；因此，所提微分方程的恰当性得到验证。

注意，我们对 $M$ 和 $N$ 项关于变量 $y$ 出现的方式并未施加任何限制。因此，恰当微分方程完全可以是非线性的。然而，它总是一阶和一次的方程。

作为恰当方程，微分方程只是表明

$$dF(y,t)=0$$

因此，其通解的形式显然为

$$F(y,t)=c$$

所以，解恰当微分方程基本上是求原函数 $F(y,t)$，并令其等于任意常数。下面，我们对方程 $M\,dy+N\,dt=0$，介绍求 $F(y,t)$ 的方法。

## 解　法

首先，因为 $M=\partial F/\partial y$，所以函数 $F$ 必定包含 $M$ 对变量 $y$ 的积分；这样，我们可以将初步结果以未确定的形式，写出如下：

$$F(y,t)=\int M\,dy+\psi(t) \tag{14.19}$$

这里，偏导数 $M$ 将仅对 $y$ 积分；即 $t$ 在积分过程中将被视为常数，正如它在 $F(y,t)$ 的偏微分从而产生 $M=\partial F/\partial y$ 的过程中被作为常数一样。[①] 因为在 $F(y,t)$ 对 $y$ 偏微分过程中，任何仅含有变

① 某些作者使用算符 $(\cdots\cdots)\partial y$ 以强调仅对 $y$ 积分。我们在这里仍使用符号 $\int(\cdots\cdots)dy$，因为几乎不存在混淆的可能性。

量 $t$ 和(或)某些常数(不含有 $y$)的相加的项会消失,因此,我们现在在积分过程中必须加以小心,以恢复这些项。这可以解释为什么在(14.19)中我们引入了一个一般项 $\psi(t)$ 的原因。尽管它并非恰好与积分常数相同,但它确实与积分常数发挥同样的作用。得到 $\int M\,dy$ 是相对容易的,但我们如何确定 $\psi(t)$项的确切形式呢?

诀窍在于利用 $N=\partial F/\partial t$。但最好还是通过例子解释具体例子。

485 **例 1** 解恰当微分方程

$$2yt\,dy + y^2 dt = 0 \qquad [(14.16)\text{ 的重写}]$$

在此方程中,有

$$M = 2yt \text{ 和 } N = y^2$$

第一步:由(14.19),我们可以首先写出初步结果

$$F(y,t) = \int 2yt\,dy + \psi(t) = y^2 t + \psi(t)$$

注意,我们已省略了积分常数,因为它能自动包含在表达式 $\psi(t)$ 中。

第二步:若我们将上述结果对 $t$ 偏微分,可以得到

$$\frac{\partial F}{\partial t} = y^2 + \psi'(t)$$

但因 $N=\partial F/\partial t$,我们可以使 $N=y^2$ 和 $\partial F/\partial t = y^2+\psi'(t)$,得到

$$\psi'(t) = 0$$

第三步:上述结果的积分得出

$$\psi(t) = \int \psi'(t)\,dt = \int 0\,dt = k$$

则我们有了 $\psi(t)$的具体形式。在本例中,$\psi(t)$恰好仅是一个常数。在更一般的情况下,它可以是 $t$ 的非常数函数。

第四步:将第一步和第三步的结果结合起来可以得到

$$F(y,t)=y^2t+k$$

则恰当微分方程的解应为 $F(y,t)=c$。但因常数 $k$ 可以纳入 $c$ 中,我们可以简单地将解写成

$$y^2t=c \quad 或 \quad y(t)=ct^{-1/2}$$

其中 $c$ 为任意常数。

**例2** 解方程 $(t+2y)dy+(y+3t^2)dt=0$。我们首先检验它是否是恰当微分方程。令 $M=t+2y, N=y+3t^2$,可求得 $\partial M/\partial t=1=\partial N/\partial y$。因此,方程通过了正合(恰当)性检验。为求其解,我们仍遵循例1的步骤。

第一步:应用(14.19)并写成

$$F(y,t)=\int(t+2y)dt+\psi(t)=yt+y^2+\psi(t)$$ 486

[常数合并于 $\psi(t)$ 中]

第二步:将此结果对 $t$ 微分,得到

$$\frac{\partial F}{\partial t}=y+\psi'(t)$$

则令其等于 $N=y+3t^2$,可求得

$$\psi'(t)=3t^2$$

第三步:将最后结果积分得到

$$\psi(t)=\int 3t^2dt=t^3 \qquad [常数可以省略]$$

第四步:将第一步和第三步的结果合并,以得到函数 $F(y,t)$ 的完备形式:

$$F(y,t)=yt+y^2+t^3$$

这意味着给定微分方程的解为

$$yt + y^2 + t^3 = c$$

读者可以验证,令此方程的全微分等于零,确实会产生给定的微分方程。

这四个步骤可以用于解任何恰当微分方程。有意思的是,甚至当给定方程不是恰当的时候,也可以应用这四个步骤。但要看到这一点,我们必须首先引入积分因子这个概念。

## 积分因子

有时,将微分方程的每一项都乘以一个特定的公因子,非恰当的微分方程也可以成为恰当微分方程。这样的因子称作**积分因子**。

**例 3** 微分方程

$$2t\,dy + y\,dt = 0$$

不是恰当的,因为它并不满足(14.18):

$$\frac{\partial M}{\partial t} = \frac{\partial}{\partial t}(2t) = 2 \neq \frac{\partial N}{\partial y} = \frac{\partial}{\partial y}(y) = 1$$

但是,如果将给定方程的每项均乘以 $y$,它便变成(14.16),从而成为恰当微分方程。因此,$y$ 是本例给出的微分方程的积分因子。

487

当可以求得一个非恰当方程的积分因子时,总可以使其成为恰当的微分方程,这样便可以应用四个解题步骤了。

## 一阶线性微分方程的解

一般的一阶线性微分方程

$$\frac{dy}{dt} + uy = w$$

按(14.17)的形式,可以表示成

$$dy + (uy - w)dt = 0 \qquad (14.20)$$

具有积分因子

$$e^{\int u dt} \equiv \exp\left(\int u \, dt\right)$$

这个形式非常不直观的积分因子可以通过如下办法“发现”：令 $I$ 为尚属未知的积分因子。以 $I$ 通乘(14.20)可以将其变为恰当微分方程：

$$\underbrace{I}_{M} dy + \underbrace{I(uy - w)}_{N} dt = 0 \qquad (14.20')$$

正合性检验表明 $\partial M/\partial t = \partial N/\partial y$。观察 $M$ 和 $N$ 表达式可知，因为 $M$ 仅由 $I$ 构成，且因 $u$ 和 $w$ 仅是 $t$ 的函数，所以，如果 $I$ 也只是 $t$ 的函数，正合性检验将简化为非常简单的条件。那时，检验变成

$$\frac{dI}{dt} = Iu \quad 或 \quad \frac{dI/dt}{I} = u$$

因此，如果 $I$ 与 $u$(或更具体地 $u(t)$)具有相同的增长率，那么，特定形式 $I = I(t)$确实成立。相应地，$I(t)$应取如下具体形式：

$$I(t) = Ae^{\int u \, dt} \qquad [参见(14.13)和(14.14)]$$

但是，很容易验证，可令常数 $A$ 等于 1 而不影响 $I(t)$满足恰当性检验的能力。因此，我们可以使用更简单的形式 $e^{\int u dt}$ 为积分因子。

将此积分因子代入(14.20′)，可以产生恰当微分方程

$$e^{\int u \, dt}dy + e^{\int u \, dt}(uy - w)dt = 0 \qquad (14.20'')$$

则此式可以通过上述四步骤方法解出。

488

第一步：首先应用(14.19)得到

$$F(y,t) = \int e^{\int u \, dt}dy + \psi(t) = ye^{\int u \, dt} + \psi(t)$$

此结果以这种简单形式出现，是因为被积函数独立于变量 $y$。

第二步：其次，将上述结果对 $t$ 微分，得到

$$\frac{\partial F}{\partial t} = yue^{\int u\,dt} + \psi'(t) \qquad [\text{链式法则}]$$

而且，因为此式可以等于 $N = e^{\int u\,dt}(uy - w)$，我们有

$$\psi'(t) = -we^{\int u\,dt}$$

第三步： 现在直接微分产生

$$\psi(t) = -\int we^{\int u\,dt}dt$$

由于未给出函数 $u = u(t)$，$w = w(t)$的具体形式，难以进行进一步的积分，而且我们必须满足于 $\psi(t)$这个非常一般的表达式。

第四步： 将此 $\psi(t)$表达式代入第一步的结果，可以求得

$$F(y,t) = ye^{\int u\,dt} - \int we^{\int u\,dt}dt$$

从而，恰当微分方程(14.20″)的通解——以及等价的，但非恰当的一阶线性微分方程(14.20)——为

$$ye^{\int u\,dt} - \int we^{\int u\,dt}dt = c$$

整理，并以 $A$ 代替(任意常数)符号 $c$，上式可以写成

$$y(t) = e^{-\int u\,dt}\left(A + \int we^{\int u\,dt}dt\right) \qquad (14.21)$$

它恰好是前面(14.15)给出的结果。

## 练习 14.4

1 验证下列微分方程是恰当的，并通过四步法解方程

(a) $2yt^3dy + 3y^2t^2dt = 0$

($b$) $3y^2t\,dy+(y^3+2t)dt=0$

($c$) $t(1+2y)dy+y(1+y)dt=0$

($d$) $\dfrac{dy}{dt}+\dfrac{2y^4t+3t^2}{4y^3t^2}=0$ [提示：首先转换成(14.17)的形式]

2　下列微分方程是恰当的吗？如果不是，尝试让 $t$、$y$ 及 $y^2$ 作为可能的积分因子。

489

($a$) $2(t^3+1)dy+3yt^2dt=0$

($b$) $4y^3t\,dy+(2y^4+3t)dt=0$

3　将四步法应用于一般恰当微分方程 $M\,dy+N\,dt=0$，推导出一个恰当微分方程通解的如下公式

$$\int M\,dy+\int N\,dt-\int\left(\frac{\partial}{\partial t}\int M\,dy\right)dt=c$$

---

## 14.5　一阶一次非线性微分方程

在**线性**微分方程中，我们不仅将导数 $dy/dt$，因变量 $y$ 限定为一次的，而且也不允许出现乘积 $y(dy/dt)$。当 $y$ 以高于一次幂形式出现时，即便方程只含有一次导数 $dy/dt$，它也是**非线性**方程。一般而言，如下形式的方程

$$f(y,t)dy+g(y,t)dt=0 \tag{14.22}$$

或

$$\frac{dy}{dt}=h(y,t) \tag{14,22'}$$

其中对 $y$ 和 $t$ 的幂数没有限制，构成了一阶一次非线性微分方程。其中某些类型的方程，用近乎常规的方法就可以相对容易地解出。我们将简单讨论三种类型。

## 恰当微分方程

我们首先讨论现在已经熟悉了的恰当微分方程。正如前面已指出的那样，$y$ 变量可在恰当方程中以更高的幂数出现，如在(14.16)——$2ytdy + y^2dt = 0$——那样[读者可将其与(14.22)相比较]。确实，从方程左边的两项中删除公共因子 $y$ 将使方程简化为线性形式，但在此情况下方程将失去其恰当性质。因此，作为恰当微分方程，必须将其视为非线性方程。

因为我们已讨论过恰当微分方程的解法，所以无须作进一步说明。

## 可分离变量

微分方程(14.22)

$$f(y,t)dy + g(y,t)dt = 0$$

490 可能恰好具有这样的方便性质：函数 $f$ 仅有变量 $y$，而 $y$ 仅含有变量 $t$，所以方程可以简化为特殊形式

$$f(y)dy + g(t)dt = 0 \tag{14.23}$$

在这种情况下，便称变量是可分离的，因为包含 $y$[合并在 $f(y)$ 中]的项在数学上可与包含 $t$ 的项[合并在 $g(t)$项中]相分离。解这种特殊类型的方程，只需简单的积分法。

**例 1** 解方程 $3y^2dy + t\,dt = 0$。首先将方程重写成

$$3y^2dy = t\,dt$$

将方程两边(均是微分)积分，并令所得结果相等，得到

$$\int 3y^2dy = \int t\,dy \qquad \text{或} \qquad y^3 + c_1 = \frac{1}{2}t^2 + c_2$$

因此,可将通解写成

$$y^3 = \frac{1}{2}t^2 + c \quad 或 \quad y(t) = (\frac{1}{2}t^2 + c)^{1/3}$$

这里值得注意的是,每一项的积分均是对不同变量进行的。正是这一点才使得可变离变量方程比较易解。

**例 2** 解方程 $2t\,dy + y\,dt = 0$。初看起来,这个微分方程似乎不属于这种类型,因为它不符合(14.23)的一般形式。具体地说,$dy$ 和 $dt$ 的系数似乎包含了"错误的"变量。然而,通过一种简单的变量,即以 $2yt(\neq 0)$ 通除各项,便可以将方程转化为可分离变量形式

$$\frac{1}{y}dy + \frac{1}{2t}dt = 0$$

根据源于例 1 的经验,可由如下步骤求解(无须首先移项):[①]

$$\int \frac{1}{y}dy + \int \frac{1}{2t}dt = c$$

所以 $\quad \ln y + \frac{1}{2}\ln t = c \quad$ 或 $\quad \ln(yt^{1/2}) = c$

因此,解为

$$yt^{1/2} = e^c = k \quad 或 \quad y(t) = kt^{-1/2}$$

491

其中 $k$ 正如符号 $c$ 和 $A$ 一样,是一个任意常数。

注意,例 2 中的方程也可以用另一种方法来解:首先通过积分因子 $y$ 将其转变成为微分方程,然后再按解恰当微分方程的办法来解它。其解我们早已在 14.4 节的例 1 中得到,它必然与刚刚用变量分离法得到的结果一致。问题在于一个给定的微分方程不仅

---

① 严格地讲,我们应将积分结果写成 $\ln|y|$ 和 $1/2\ \ln|t|$。如果能把 $y$ 和 $t$ 假设为正,(在多数经济问题中,这样假设是合适的)那么,就可以得出文中给出的结果。

只有一种解法，因此，人们可以选择所使用的方法。在其它情况下，不适合特殊解法的微分方程，经适当变换，也可能适用这种解法。

### 可简化为线性的方程

如果微分方程 $dy/dt = h(y, t)$ 恰好取特定的非线性形式：

$$\frac{dy}{dt} + Ry = Ty^m \tag{14.24}$$

其中 $R$ 和 $T$ 是两个 $t$ 的函数，$m$ 是不为 0 和 1 的任意数（若 $m = 0$ 或 $m = 1$ 会如何？），则此方程——称作伯努利方程——总可以简化为线性微分方程，并且可按线性微分方程的解法去解它。

转化为线性方程的步骤是相对简单的。首先以 $y^m$ 除（14.24），得到

$$y^{-m}\frac{dy}{dt} + Ry^{1-m} = T$$

若我们采用如下简写变量

$$z = y^{1-m}\left[\text{则}\frac{dz}{dt} = \frac{dz}{dy}\frac{dy}{dt} = (1-m)y^{-m}\frac{dy}{dt}\right]$$

则上述方程可以写成

$$\frac{1}{1-m}\frac{dz}{dt} + Rz = T$$

进而，在以 $(1-m)dt$ 通乘并重排方程后，我们可将方程转化为

$$dz + [(1-m)Rz - (1-m)T]dt = 0 \tag{14.24'}$$

可以看出，这是与（14.20）形式一致的一阶线性微分方程，其中变量 $z$ 取代了变量 $y$。

显然，我们可以应用公式（14.21）去求其解 $z(t)$。最后，再用

逆代换，将 $z$ 变换成 $y$。

**例 3** 解方程 $dy/dt+ty=3ty^2$。这是一个伯努利方程，其中 $m=2$（给出 $z=y^{1-m}=y^{-1}$），$R=t$，$T=3t$。因此，由(14.24′)，492 我们可以写出如下线性微分方程

$$dz+(-tz+3t)dt=0$$

应用公式(14.21)，可求得解为

$$z(t)=A\exp(\frac{1}{2}t^2)+3$$

（作为练习，请读者自己写出得到此解的具体步骤。）

因为我们原来感兴趣的是得到解 $y(t)$ 而非 $z(t)$，所以，我们必须用方程 $z=y^{-1}$ 或 $y=z^{-1}$ 进行逆变换。因此，取 $z(t)$ 的倒数，得到所求的解

$$y(t)=\frac{1}{A\exp(\frac{1}{2}t^2)+3}$$

因为存在任意常数 $A$，所以它是一个通解。

**例 4** 解方程 $dy/dt+(1/t)y=y^3$。这里我们有 $m=3$（因此 $z=y^{-2}$），$R=1/t$，$T=1$。因此，方程可以线性化为

$$dz+(\frac{-2}{t}z+2)dt=0$$

读者可以验证，根据公式(14.21)，此微分方程的解为

$$z(t)=At^2+2t$$

由此可知，通过逆变换 $y=2^{-1/2}$，原变量的通解可以写成

$$y(t)=(At^2+2t)^{-1/2}$$

作为一个练习，读者可通过微分检验上述两个例子解的正确性。

## 练习 14.5

1 确定下列每个方程(1)是否是可分离变量方程;(2)是否是线性的或可线性化的。

(a) $2t\,dy + 2y\,dt = 0$　　　(c) $\frac{dy}{dt} = -\frac{t}{y}$

(b) $\frac{y}{y+t}dy + \frac{2t}{y+t}dt = 0$　　　(d) $\frac{dy}{dt} = 3y^2t$

2 取 $y$ 和 $b$ 为正,运用变量分离法解练习 14.5-1 中的($a$)和($b$)题。

3 将练习 14.5-1 中的($c$)题作为可分离变量方程和伯努利方程,分别

493 解之。

4 将练习 14.5-1 中的($d$)题作为可分离变量方程和伯努利方程,分别解之。

5 通过证明例 4 的中间解 $z(t) = At^2 + 2t$ 的导数 $dz/dt$ 与线性化的微分方程相一致,来验证中间解 $z(t) = At^2 + 2t$ 的正确性。

# 14.6 定性图解法

前面讨论的几种非线性微分方程(恰当微分方程、可分离变量方程及伯努利方程)均可定量解出。亦即在每种情况下,我们都求得了时间路径方程 $y(t)$;对于每个 $t$ 值,该方程确定了具体对应的 $y$ 值。

有时,我们难以求出给定微分方程的定量解。然而,在这种情况下,通过直接观察微分方程或者分析其图形,有可能确定时间路径的定性性质[主要指 $y(t)$是否收敛]。而且,即使能够得到定量解,但如果我们主要关心的是时间路径的定性性质,那么,我们仍

可使用定性分析方法。

## 相位图

给定一般形式的一阶微分方程

$$\frac{dy}{dt} = f(y)$$

它是变量 $y$ 的线性或非线性方程。在图 14.3 中，我们绘出了 $dy/$

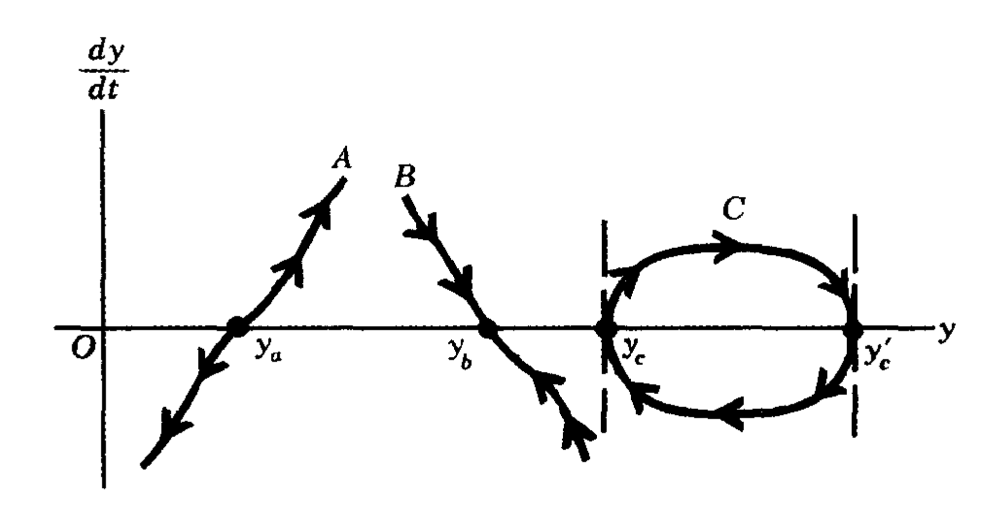



图 14.3

$dt$ 对 $y$的曲线。只要 $dy/dt$ 仅是 $y$ 的函数，便可以作这种几何表示，并称其为相位图；而表示函数 $f$ 的图形，称作相位线。（时间变量 $t$ 不是函数 $f$ 的分离变量的那种形式的微分方程，称作自主 494 微分方程。）一旦相位线已知，其图形将给出关于时间路径 $y(t)$的重要信息。这种定性的信息概括在如下两个结论中：

1. 在横轴上方任意点（其中 $dy/dt>0$），$y$ 必随时间而递增，而且对 $y$ 轴而言，$y$ 必由左向右移动。同理，横轴下方的任意点必与 $y$ 向左移动相联系，因为 $dy/dt$ 为负意味着 $y$ 随时间而递减。这些方向的趋势解释了图 14.4 中的相位图中的箭头之所以那样画的原因。在横轴上方，箭头均指向右方——东北、东南或正东，

具体指向视情况而定。在 $y$ 轴下方，则情况恰恰相反。而且，这些结果与 $y$ 的代数符号无关；既便相位线 $A$（或其它相位线）移至纵轴左边，对箭头的方向也无影响。

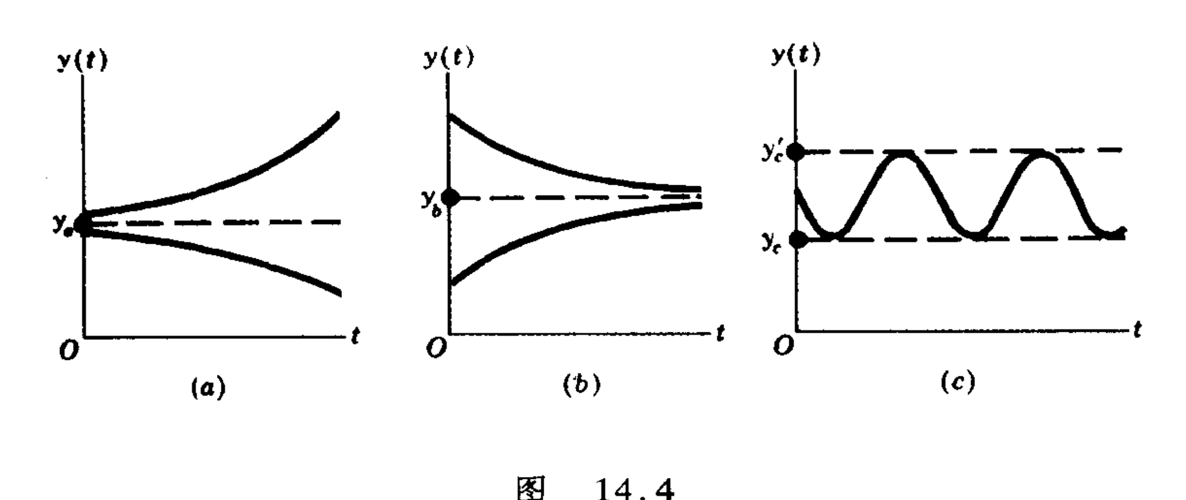



图 14.4

2. 如果 $y$ 的均衡水平（瞬时意义上的均衡）存在的话，也仅能存在于横轴上，此处 $dy/dt=0$（$y$ 对时间稳定）。因此，为求出均衡水平，只须考察相位线与横轴的交点。① 另一方面，为检验均衡的动态稳定性，不管 $y$ 的初始状态如何，我们还应检验相位线在所说的交点是否总是指向均衡位置。

## 时间路径的类型

在上述概括的基础上，由图 14.3 所示的相位线中，我们可以观察到三种时间路径的类型。

相位线 $A$ 在点 $y_a$ 有个均衡。但在这点的上部和下部，箭头总是指向偏离均衡的方向。因此，尽管如果恰好 $y(0)=y_a$，可以达到均衡；但更一般的 $y(0)\neq y_a$ 的情况，将导致 $y$ 不断增加[若 $y$

① 但并非所有交点都代表均衡状态。在讨论图 14.3 中的相位线 $C$ 时，将会看到这一点。

$(0)>y_a$]，或不断减少[若 $y(0)<y_a$]。此外，在这种情况下，$y$ 对 $y_a$ 的偏离趋向于以递增的速度增长，因为当我们在相位线上沿着箭头移动时，将会进一步偏离 $y$ 轴，因而遇到 $dy/dt$ 的数值也不断增加的情况。因此，相位线 $A$ 所意指的时间路径 $y(t)$ 可用图 14.4$a$ 所示的曲线来表示。其中的图形是以 $y$ 对 $t$ 绘成的，而非以 $dy/dt$ 对 $y$ 绘成的。均衡 $y_a$ 为动态不稳定。

相反，相位线 $B$ 在 $y_b$ 存在稳定均衡。如果 $y(0)=y_b$，可以立即达到均衡。但相位线 $B$ 的重要特征在于，即使 $y(0)\neq y_b$，然而沿着相位线的移动将使 $y$ 趋向于 $y_b$ 水平。因而，对应于这种相位 495 线的时间路径的形式如图 14.4$b$；它使我们联想到动态市场模型。

上述讨论表明：一般而言，相位线在其交点处的斜率是决定均衡动态稳定性或者时间路径收敛性的关键。(有限)正斜率，比如在点 $y_a$，会产生动态不稳定性；(有限)负斜率，比如在点 $y_b$，意味着动态稳定性。

甚至无需绘出微分方程的相位线，上述概括也能帮助我们推断其定性结论。比如，以(14.4)中的线性微分方程为例，

$$\frac{dy}{dt}+ay=b \quad \text{或} \quad \frac{dy}{dt}=-ay+b$$

因为相位线显然具有不变斜率 $-a$（这里假定为非零），我们可以立即推断出(毋需绘出相位线)

$$a \gtrless 0 \Leftrightarrow y(t)\left\{\begin{matrix}\text{收敛于}\\ \text{偏离于}\end{matrix}\right\}\text{均衡}$$

正如我们可以预期的那样，此结果与此方程的定量解

$$y(t)=\left[y(0)-\frac{b}{a}\right]e^{-at}+\frac{b}{a} \qquad [\text{由}(14.5')]$$

所揭示的结论完全一致。我们已经知道，从非均衡状态出发，$y$

$(t)$的收敛性取决于当$t \to \infty$时，$e^{-at} \to 0$。但此结果当且仅当$a>0$时才会发生。若$a<0$，则当$t \to \infty$时，$e^{-at} \to \infty$，因而$y(t)$不能收敛。因此，无论是定量分析还是定性分析，我们的结论是一致的，相同的。

相位线$C$尚有待于讨论。它是与横轴交叉的闭合环状曲线；它不是一个函数，只表明$dy/dt$和$y$之间的关系。[1] 在此情况下，出现的一个重要的新因素是时间路径具有周期波动的可能性。从绘出的相位线$C$的图形，我们可以看到$y$在两个值，即$y_c$及$y_c'$之间不断波动。为了产生周期波动，环状曲线须以这种方式与横轴交叉，从而使$dy/dt$交替出现正值和负值。此外，在两个交点$y_c$和$y_c'$，相位线的斜率应为无穷大，否则交点就类似$y_a$或$y_b$，在$y_a$和$y_b$都不允许箭号连续移动。对应于环形相位线的时间路径已在图14.c中作了说明。注意，只要$y(t)$达到上界$y_c'$或下界$y_c$，总有$dy/dt=0$(局部极值)；但毫无疑问，这些值不表示$y$的均衡值。根据图14.3，这意味着相位线与$y$轴的所有交点并不都是均衡位置。

496

总之，为研究均衡的动态稳定性(或时间路径的收敛性)，我们既可以求出时间路径本身，也可以从其相位线简单推断。我们将用索洛增长模型说明后一方法的应用。

① 此关系产生于二次微分方程$(dy/dt)^2=f(y)$。

## 练习 14.6

---

1 绘出下列每一方程的相位线,并讨论其定性含义:

(a) $\frac{dy}{dt}=y-7$　　(c) $\frac{dy}{dt}=4-\frac{y}{2}$

(b) $\frac{dy}{dt}=1-5y$　　(d) $\frac{dy}{dt}=9y-11$

2 绘出下列每一方程的相位线,并对其进行解释:

(a) $\frac{dy}{dt}=(y+1)^2-16 \qquad (y\geq 0)$

(b) $\frac{dy}{dt}=\frac{1}{2}y-y^2 \qquad (y\geq 0)$

3 给定 $dy/dt=(y-3)(y-5)=y^2-8y+15$

(a) 请推断:$y$ 的可能的均衡水平有两个,一个是在 $y=3$,另一个在 $y=5$。

(b) 分别求$\frac{d}{dy}\left(\frac{dy}{dt}\right)$在 $y=3$ 和 $y=5$ 的符号。由此你能推断出何结论?

---

# 14.7 索洛增长模型

索洛教授的增长模型[①] 的意义之一在于证明了多马模型的刃锋增长路径主要是模型所采纳的特殊生产函数假设的结果,还 497 证明了在其它情况下,对敏感均衡的需要也许不会产生。

## 框　架

在多马模型中,产出明确表述为仅是资本的函数;$k=\rho K$(生

① 罗伯特·M.索洛:"对经济增长理论的贡献",载《经济学季刊》1956 年 2 月,65—94页。

产能力或潜在产出，是资本存量的常数倍数)。生产函数中没有劳动投入表明这样一种含义：劳动与资本总按固定比例组合，因而只需考察那些生产函数中的一种。与此相反，索洛试图分析资本和劳动按可变比例组合的情况。因此，其生产函数的形式为

$$Q = f(K, L) \quad (K, L > 0)$$

其中 $Q$ 为产出(不含折旧)，$L$ 为劳动，$K$ 为资本——所有变量均在宏观意义上所用。假设 $f_K$ 为 $f_L$ 为正(边际产量为正)，$f_{KK}$ 和 $f_{LL}$ 为负(每种投入的边际收益递减)。进而，假设生产函数 $f$ 为线性齐次函数(不变规模收益)。所以，能够写出

$$Q = Lf\left(\frac{K}{L}, 1\right) = L\phi(k) \qquad \text{在此 } k \equiv \frac{K}{L} \tag{14.25}$$

考虑到假设的 $f_K$ 和 $f_{KK}$ 的符号，新引进的 $\phi$ 函数(注意，它只有单一变数 $k$)必具有一阶导数为正，二阶导数为负的特征。为验证此结论，回顾(12.49)可知

$$f_K \equiv \text{MPP}_K = \phi'(k)$$

因此，$f_K > 0$ 必然意味着 $\phi'(k) > 0$。则因

$$f_{KK} = \frac{\partial}{\partial K}\phi'(k) = \frac{d\phi'(k)}{dk}\frac{\partial k}{\partial K} = \phi''(k)\frac{1}{L} \qquad [\text{见}(12.48)]$$

假设 $f_{KK} < 0$ 直接引出结论 $\phi''(k) < 0$。因此，$\phi$ 函数——按照(12.46)，它给出了每一资本劳动比率的 $\text{APP}_L$——是一个按递减速率随 $k$ 递增的函数。

给定 $Q$ 取决于 $K$ 和 $L$，现在有必要规定后两个变量本身是如何决定的。索洛的假设是：

$$\dot{K}\left(\equiv \frac{dK}{dt}\right) = sQ \qquad [Q \text{ 的固定比例用于投资}] \tag{14.26}$$

$$\frac{\dot{L}}{L}\left(\equiv \frac{dL/dt}{L}\right) = \lambda\ (\lambda > 0) \qquad [\text{劳动力呈指数增长}] \quad (14.27)$$

符号 $S$ 表示不变边际储蓄倾向,λ 表示(不变)劳动增长率。注意这些假设的动态性质;它们不但规定了 $K$ 和 $L$ 的水平是如何决定的,而且也设定了其变化率是如何决定的。

方程(14.25)、(14.26)、(14.27)构成了一个完整的模型。为 498 解此模型,我们需要首先将其变换成一个单变量方程。将(4.25)代入(4.26),得到

$$\dot{K} = sL\phi(k) \quad (14.28)$$

因 $k \equiv K/L$, $K \equiv kL$,所以将后一恒等式微分,可以得到 $\dot{K}$ 的另一表达式:

$$\begin{aligned}\dot{K} &= L\dot{k} + k\dot{L} \quad [\text{积的微分法则}] \\ &= L\dot{k} + k\lambda L \qquad [\text{由}(14.27)] \end{aligned} \quad (14.29)$$

令(14.29)等于(14.28),并消去公共因子 $L$,出现结果

$$\dot{k} = s\phi(k) - \lambda k \quad (14.30)$$

这个以 $k$ 为变量,具有两个参数 $S$ 和 λ 的微分方程,是索洛增长模型的基本方程。

## 定性图解分析

因为(14.30)是以一般函数形式表述的,不能得到具体的定量解。但我们可以对其进行定性分析。为此,我们需要给出以 $\dot{k}$ 为纵轴,以 $k$ 为横轴的相位线。

因(14.30)的右边包含两项,所以,首先将其绘成两条各自独

立的曲线。$\lambda k$ 项，是 $k$ 的线性函数，在图 14.5$a$ 中显然是一条向右上方倾斜的直线，其纵截距为零，斜率为 $\lambda$。$s\phi(k)$ 项，可以绘成一条以递减速率递增的曲线。因为 $s\phi(k)$ 只是 $\phi(k)$ 的不变分数倍。若我们将 $K$ 视为必不可少的生产要素，$s\phi(k)$ 曲线必须起自原点；这是因为若 $K=0$，因而 $k=0$，必然有 $Q$ 为零，$\phi(k)$ 及 $s\phi(k)$ 也为零。曲线实际绘出的方式也反映了这样一个隐含的假设：存在着一个 $k$ 值的集合，使得 $s\phi(k)$ 超过 $\lambda k$，从而使得两条曲线在 $k$ 的某个正值，比如 $\bar{k}$，相交。

基于上述两条曲线，对于每个 $k$ 值，$\dot{k}$ 值可以由两条曲线间的纵向距离来度量。像图 14.5 那样，对 $k$ 绘出 $\dot{k}$ 的值，便会产生我们需要的相位线。注意，当资本劳动比率为 $\bar{k}$ 时图 $a$ 中的两条曲线相交，图 $b$ 中的相位线必在 $\bar{k}$ 与横轴相交。这标志着 $\bar{k}$ 为（瞬时）均衡资本劳动比率。

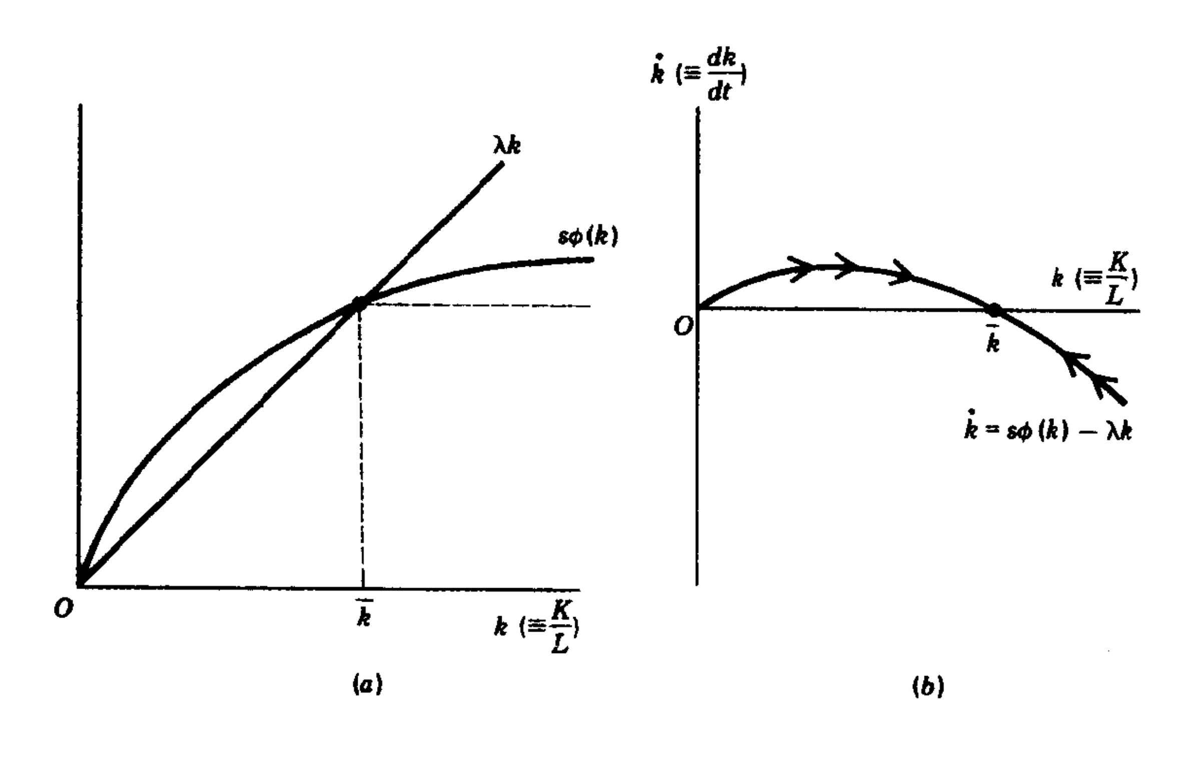



图　14.5

由于相位线在 $\bar{k}$ 的斜率为负,均衡显然是稳定均衡。给定任意正的初始值 $k$,模型的动态运动必然使之收敛于均衡水平 $\bar{k}$。重要的一点是,一旦达到均衡,则由定义,资本劳动比率不随时间而变化,其后,资本必随劳动以同一比例 $\lambda$ 迅速增长。这又意味着净投资必定以比率 $\lambda$(参见练习 14.7-2)增长。但要注意,这里使用"必定"一词并不含有"要求"的含义,而是指"自动"的意思。因此,索洛模型用来说明,给定劳动增长率 $\lambda$,无需按照多马模型的灵敏平衡方式,该经济体本身也会达到稳定增长的均衡状态,在这种均衡状态中,投资同 $K$ 和 $L$ 一样,也以比率 $\lambda$ 增长。而且,为满足(14.25),$Q$ 必须也以同样的比率增长,因为当资本劳动比率在 $\bar{k}$ 水平保持不变时,$\phi(k)$ 是一个常数。这种所有有关变量均按同一比率增长的状态,称作**稳定状态**(*Steady state*)——这是静止状态(*Stationary state*)概念的推广。(在静止状态中,相关变量保持不变,换言之,其增长率为零)。 499

注意,在上述分析中,为方便起见,假定生产函数不随时间而变化。但如果技术状况得到改善,那么,生产函数必须作适当修正。比如,可以将其写成这种形式

$$Q = T(t)f(K, L) \qquad \left(\frac{dT}{dt} > 0\right)$$

其中 $T$ 是技术水平的度量;它是时间的增函数。由于相乘的项 $T(t)$ 随时间而递增,$K$ 和 $L$ 不变也会使未来某一时间的产出大于现在的产出。在此情况下,图 14.5 中的 $s\phi(k)$ 曲线将持续向上移动,使之与 $\lambda k$ 射线的交点不断上升,并使 $\bar{k}$ 值不断增大。因此,随着技术的进步,在连续稳定状态中,该经济体中的每个工人可利用的资本设备的数量日益增多,从而使生产力日益提高,这一切都 500

将是可能的。

### 定量解释

由于模型中存在一般函数 $\phi(k)$，所以上述分析必然是定性的。但是，如果我们设定一种具体形式的生产函数，比如线性柯布—道格拉斯生产函数，那么，便可以求得定量解。

我们将生产函数写成

$$Q = K^{\alpha}L^{1-\alpha} = L\left(\frac{K}{L}\right)^{\alpha} = Lk^{\alpha}$$

从而 $\phi(k) = k^{\alpha}$。则(14.30)变成

$$\dot{k} = sk^{\alpha} - \lambda k \qquad 或 \qquad \dot{k} + \lambda k = sk^{\alpha}$$

它是以 $k$ 为变量[见(14.24)]，有 $R = \lambda, T = s, m = \alpha$ 的伯努利方程。令 $z = k^{1-\alpha}$，我们得到其线性形式：

$$dz + [(1-\alpha)\lambda z - (1-\alpha)s]dt = 0$$

$$或 \qquad \frac{dz}{dt} + \underbrace{(1-\alpha)\lambda}_{a} z = \underbrace{(1-\alpha)s}_{b}$$

这是一个具有常系数 $a$ 和常数项 $b$ 的线性微分方程。因此，由公式(14.5′)，我们有

$$z(t) = \left[z(0) - \frac{s}{\lambda}\right]e^{-(1-\alpha)\lambda t} + \frac{s}{\lambda}$$

代入 $z = k^{1-\alpha}$ 便会得到最终解

$$k^{1-\alpha} = \left[k(0)^{1-\alpha} - \frac{s}{\lambda}\right]e^{-(1-\alpha)\lambda t} + \frac{s}{\lambda}$$

其中 $k(0)$ 为资本劳动比率 $k$ 的初始值。

此解确定 $k$ 的时间路径。回顾一下，$(1-\alpha)$ 和 $\lambda$ 均为正，可以看到当 $k \to \infty$ 时，指数式将趋于零，因而

$$k^{1-\alpha} \to \frac{s}{\lambda} \quad \text{或} \quad k \to \left(\frac{s}{\lambda}\right)^{1/(1-\alpha)} \quad \text{当} \quad t \to \infty$$

于是，资本劳动比率将趋近于一个成为其均衡值的常数。此均衡或者稳定状态值$(s/\lambda)^{1/(1-\alpha)}$与边际储蓄倾向 $s$ 成正比，与资本劳动比率成反比。

## 练习 14.7

---

1 以 $k$ 通除(14.30)，按照 $k$、$K$、$L$ 的增长率解释所得到的方程。

2 证明：如果资本按比率 $\lambda$ 增长（即 $K = Ae^{\lambda t}$），净投资 $I$ 必然也按比率 $\lambda$ 增长。 501

3 如果索洛模型中的生产函数为 $Q = T(t)f(K,L)$，其中函数 $f$ 为线性齐次的，$T(t)$按比率 $\rho$ 增长，$L$ 按比率 $\lambda$ 增长，那么，稳定状态下 $Q$ 的增长率必然为多少？

4 索洛模型中的原投入变量为 $K$ 和 $L$，但基本方程(14.30)只集中于资本劳动比率 $k$，模型中的哪个（些）假设与这种变量转换有关（或使其可能）？试解释。

5 绘出下列每个方程的相位图，并讨论时间路径 $y(t)$的定性性质：

$(a)\ \dot{y} = 3 - y - \ln y \quad (b)\ \dot{y} = e^{y} - (y+2)$

---

# 第 15 章　高阶微分方程

502　在上一章，我们讨论了解一阶微分方程的方法。在一阶微分方程中，不存在阶数高于 1 的导数或微分。但有时，设定的模型可能包括二阶甚至更高阶的导数。比如，我们可能得到一个描述收入 $Y$“变化率的变化率”：

$$\frac{d^2Y}{dt^2} = kY$$

我们需要由其求得 $Y$ 的时间路径。在此情况下，给定函数构成了二阶微分方程，求时间路径 $Y(t)$ 即解二阶微分方程。本章关心的是这样的高阶微分方程的解法及其经济应用，但我们仍只讨论线性的情况。

一类简单的 $n$ 阶线性微分方程的形式如下：

503
$$\frac{d^ny}{dt^n} + a_1\frac{d^{n-1}y}{dt^{n-1}} + \cdots + a_{n-1}\frac{dy}{dt} + a_ny = b \qquad (15.1)$$

或用另一种符号表示

$$y^{(n)}(t) + a_1y^{(n-1)}(t) + \cdots + a_{n-1}y'(t) + a_ny = b \qquad (15.1')$$

因为方程中的第 $n$ 阶导数(左边第一项)是最高阶导数，所以它是 $n$ 阶导数。它还是线性方程，因为因变量 $y$ 及所有导数都是一次的，而且，不出现 $y$ 和其任意阶导数相乘的乘积项。此外，读者还应注意，这个微分方程特征在于常系数(各个 $a$)和常数项($b$)。系

数为常数是本章始终保持的假设。但常数项 $b$，我们只是第一步采纳这一假设，在 15.5 节，我们将放弃这一假设，而采用可变项。

## 15.1 具有常系数和常数项的二阶线性微分方程

为便于教学，我们首先讨论二阶线性微分方程的解法。有关微分方程就是这样一种简单形式

$$y''(t)+a_1y'(t)+a_2y=b \tag{15.2}$$

其中 $a_1,a_2$ 及 $b$ 均为常数。若 $b$ 恒为零，则得到齐次方程，但若 $b$ 为非零常数，则方程为非齐次方程。我们的讨论是在(15.2)为非齐次的假设下进行的；在解非齐次方程(15.2)时，自然会出现一个副产品，即齐次方程的解。

在这方面，我们回顾一下在 14.1 节引入且在这里同样可以应用的命题：如果 $y_c$ 是余函数，亦即简化方程(15.2)的通解(具有任意常数)，如果 $y_p$ 是特解，即完备方程(15.2)的任意特解(不含任意常数)，则 $y(t)=y_c+y_p$ 是完备方程的通解。正如前面所解释的那样，$y_p$ 部分使我们得到了该项瞬时意义上变量 $y$ 的均衡值，而 $y_c$ 则揭示了在每一时点，时间路径 $y(t)$ 与均衡的偏离。

### 特别积分

对于常系数和常数项的情况，特别积分是相对容易求出的。因为特别积分可以是(15.2)的任何一个解，即满足此非齐次方程的任意 $y$ 值，所以我们总可以试用最简单的函数形式，即 $y=$ 常数。如果 $y=$ 常数，由此得到

$$y'(t) = y''(t) = 0$$

504 从而(15.2)实际变成 $a_2 y = b$，解为 $y = b/a_2$。因此，所求的特别积分为

$$y_p = \frac{b}{a_2} \quad (a_2 \neq 0) \tag{15.3}$$

因为求 $y_p$ 值的过程中涉及条件 $y(t)=0$，所以，将 $y_p$ 值视为瞬时均衡，其合理性是不言自明的。

**例 1** 求方程

$$y''(t) + y'(t) - 2y = -10$$

的特别积分。这里，相关的系数是 $a_2 = -2, b = -10$。因此，特别积分为 $y_p = -10/(-2) = 5$。

如果 $a_2 = 0$，从而 $b/a_2$ 未有定义，情况会怎么样？在此情况下，因为 $y_p$ 的常数解不成立，所以必须尝试解的非常数形式。取最简单的可能形式，可令 $y = kt$。因为 $a_2 = 0$，现在微分方程为

$$y''(t) + a_1 y'(t) = b$$

但若 $y = kt$，这意味着 $y'(t) = k$，$y''(t) = 0$，此方程简化为 $a_1 k = b$。这确定了 $k$ 值为 $b/a_1$，因而给出特定积分

$$y_p = \frac{b}{a_1} t \quad (a_2 = 0; a_1 \neq 0) \tag{15.3'}$$

在此情况下，由于 $y_p$ 是时间的非常数函数，所以，我们将其视为移动均衡。

**例 2** 求方程 $y''(t) + y'(t) = -10$。这里，我们有 $a_2 = 0$，$a_1 = 1$ 及 $b = -10$。因此，由(15.3′)，我们可以写出

$$y_p = -10t$$

若 $a_1$ 恰好也为零，则解的形式 $y = kt$ 也不成立，因为 $bt/a$，

现在也未有定义。那么，我们应尝试形式为 $y=kt^2$ 的解。由于 $a_1=a_2=0$，微分方程简化为一种极其简单的形式

$$y''(t)=b$$

若 $y=kt^2$，有 $y'(t)=2kt$，$y''(t)=2k$，可将微分方程写成 $2k=b$。因此求得 $k=b/2$，且特别积分为

$$y_p=\frac{b}{2}t^2 \quad (a_1=a_2=0) \tag{15.3''}$$

此特别积分所表示的均衡也是移动均衡。

**例3** 求 $y''(t)=-10$ 的 $y_p$。因为系数为 $a_1=a_2=0$，$b=-10$，公式(15.3″)适用。所以所求答案为 $y_p=-5t^2$。

505

## 余函数

(15.2)的余函数定义为其简化(齐次)方程

$$y''(t)+a_1y'(t)+a_2y=0 \tag{15.4}$$

的通解。这就是为什么我们表明齐次函数的解总是解完备方程过程的**副产品**的原因。

尽管我们以前从未接触到这样的方程，但求一阶微分方程余函数的经验可以给我们提供有益的启示。由解(14.3)、(14.3′)、(14.5)及(14.5′)可知，指数表达式 $Ae^{rt}$ 在常系数一阶微分方程的余函数中起着非常重要的作用。那么，为什么不在二阶微分方程中也试用 $y=Ae^{rt}$ 这种解的形式呢？

若我们采用试探解 $y=Ae^{rt}$，则必有 $y$ 的导数

$$y'(t)=rAe^{rt} \quad 和 \quad y''(t)=r^2Ae^{rt}$$

在 $y$，$y'(t)$，$y''(t)$ 表达式的基础上，可将微分方程(15.4)变换为

$$Ae^{rt}(r^2+a_1r+a_2)=0 \tag{15.4'}$$

只要我们选择的那些 $A$ 和 $r$ 值满足(15.4′),则试探解 $y=Ae^{rt}$ 应成立。这意味着我们或者令 $A=0$,或者看到 $r$ 满足方程

$$r^2+a_1r+a_2=0 \tag{15.4″}$$

但因为任意常数 $A$ 可以运用问题的初始条件来确定,所以我们不能简单地随意令 $A=0$。因此,有必要求满足(15.4″)的 $r$ 值。

方程(15.4″)被称作齐次方程(15.4)或完备方程(15.2)的**特征方程**(或辅助方程)。因为它是 $r$ 的二次方程,所以产生如下两个根(解)(这里称其为**特征根**):①

$$r_1,\ r_2=\frac{-a_1\pm\sqrt{a_1^2-4a_2}}{2} \tag{15.5}$$

这两个根之间具有一种简单而又重要的关系,它可以作为检验我506们计算结果的一种方法:两个根的**和**总是等于 $-a_1$,而其**积**总是等于 $a_2$。此结论的证明很简单:

$$r_1+r_2=\frac{-a_1+\sqrt{a_1^2-4a_2}}{2}+\frac{-a_1-\sqrt{a_1^2-4a_2}}{2}=\frac{-2a_1}{2}=-a_1 \tag{15.6}$$

$$r_1r_2=\frac{(-a_1)^2-(a_1^2-4a_2)}{4}=\frac{4a_2}{4}=a_2$$

这两个根是我们可赋予解 $y=Ae^{rt}$ 中的 $r$ 的仅有的值。但这实际上意味着存在着**两个**成立的解,即

$$y_1=A_1e^{r_1t} \quad 和 \quad y_2=A_2e^{r_2t}$$

其中 $A_1$ 和 $A_2$ 为两个任意常数,$r_1,r_2$ 为由(15.5)求得的特征

---

① 注意,二次方程(15.4)是标准化形式。$r^2$ 的系数是 1。在应用公式(15.5)以求微分方程特征根时,我们必须首先确定特征方程确实是标准化形式。

根。但因我们仅需一个通解，所以有一个解似乎是多余的。我们现在有两种选择：(1)随机地选择 $y_1$ 或 $y_2$；(2) 将其以某种方式合并。

第一种方式虽简单，但却难以接受。因为在 $y_1$ 或 $y_2$ 中仅有一个任意常数，但要作为二阶微分方程的通解，式中必须含有两个任意常数。这个要求产生于如下事实：在由函数 $y(t)$ 产生二阶导数 $y''(t)$ 的过程中，我们在两轮微分过程中"失去"两个常数，所以由二阶微分方程回复到原函数 $y(t)$ 时，应恢复两个常数。这使我们只能选择 $y_1$ 和 $y_2$ 组合的形式，从而使常数 $A_1$ 和 $A_2$ 均包含在内。所以，我们可以简单地取 $y_1+y_2$ 作为(15.4)的通解。下面，我们来证明：如果 $y_1$ 和 $y_2$ 分别满足(15.4)，则和$(y_1+y_2)$也将满足(15.4)。若 $y_1$ 和 $y_2$ 确实是(15.4)的解，则将每个解代入(15.4)，必然发现如下两个方程成立：

$$y''_1(t)+a_1y'_1(t)+a_2y_1=0$$

$$y''_2(t)+a_1y'_2(t)+a_2y_2=0$$

将两个方程相加，求得

$$\underbrace{[y''_1(t)+y''_2(t)]}_{=\frac{d^2}{dt^2}(y_1+y_2)}+\underbrace{a_1[y'_1(t)+y'_2(t)]}_{=\frac{d}{dt}(y_1+y_2)}+a_2(y_1+y_2)=0$$

因此，像 $y_1$ 或 $y_2$ 一样，和$(y_1+y_2)$也满足(15.4)。相应地，齐次方程(15.4)的通解，或完备方程(15.2)的余函数一般可写成 $y_c=y_1+y_2$。

507

对特征根公式(15.5)的更深入考察表明，就 $r_1$ 和 $r_2$ 的值而论，会产生三种可能的情况。在有些情况下，有必要对结果 $y_c=$

$y_1+y_2$ 进行修改。

**第一种情况(不同的实根)** 当 $a_1^2>4a_2$ 时,(15.5)中的平方根为实数,两个根 $r_1$ 和 $r_2$ 取不同的实数值,因为 $r_1$ 等于 $-a_1$ 加上平方根,而 $r_2$ 等于 $-a_1$ 减去平方根。在此情况下,我们的确能写成

$$y_c=y_1+y_2=A_1e^{r_1t}+A_2e^{r_2t}(r_1\neq r_2) \qquad (15.7)$$

因为两个根不同,因而两个指数式必定线性无关(没有一个根是另一个的倍数);因而 $A_1$ 和 $A_2$ 总是作为独立的整体。我们便得到了所需要的两个常数。

**例4** 解微分方程

$$y''(t)+y'(t)-2y=-10$$

此方程的特别积分早已在例1中求出,为 $y_p=5$。现在我们求其余函数。因为方程的系数为 $a_1=1, a_2=-2$,由(15.5),特征根为

$$r_1, r_2=\frac{-1\pm\sqrt{1+8}}{2}=\frac{-1\pm 3}{2}=1, -2$$

(检验: $r_1+r_2=-1=-a_1$; $r_1r_2=-2=a_2$。)因为根为不同的实根,所以余函数为 $y_c=A_1e^t+A_2e^{-2t}$。因此,通解可以写成

$$y(t)=y_c+y_p=A_1e^t+A_2e^{-2t}+5 \qquad (15.8)$$

为确定常数 $A_1$ 和 $A_2$,现在需要两个初始条件。令这两个条件为 $y(0)=12$, $y'(0)=-2$。即当 $t=0$ 时,$y(t)$ 和 $y'(t)$ 分别为12和 $-2$。令(15.8)中的 $t=0$,求得

$$y(0)=A_1+A_2+5$$

将(15.8)对 $t$ 微分,然后令导数中的 $t=0$,可求得

$$y'(t)=A_1e^t-2A_2e^{-2t} \quad 和 \quad y'(0)=A_1-2A_2$$

因此,为满足两个初始条件,必须令 $y(0)=12$, $y'(0)=-2$,得到

如下两个联立方程

$$A_1 + \quad A_2 = 7$$

$$A_1 - 2A_2 = -2$$

其解为 $A_1 = 4, A_2 = 3$。因此，微分方程的定解为

$$y(t) = 4e^t + 3e^{-2t} + 5 \qquad (15.8')$$ 508

同以前一样，我们可以通过微分检验此解的正确性。(15.8′)的一阶和二阶导数为

$$y'(t) = 4e^t - 6e^{-2t} \quad \text{和} \quad y''(t) = 4e^t + 12e^{-2t}$$

当将它们与(15.8′)一起代入给定微分方程时，得到等式 $-10 = -10$。因此解是正确的。读者可以轻松地验证，(15.8′)也满足两个初始条件。

**第二种情况(重实根)** 当微分方程中的系数使得 $a_1^2 = 4a_2$，(15.5)中的特征根等于零，两个特征根取相等的值

$$r(= r_1 = r_2) = -\frac{a_1}{2}$$

这样的根称为**重根**，或**多重根**(在这里为 2 重根)。

如果我们力图将余函数写成 $y_c = y_1 + y_2$，则在此情况下，两项合作为一项：

$$y_c = A_1e^{rt} + A_2e^{rt} = (A_1 + A_2)e^{rt} = A_3e^{rt}$$

从而只含有一个常数。这使我们难以从二阶微分方程回复到其原函数。唯一的出路是求出和的另一个适当的项——它能满足(15.4)并与 $A_3e^{rt}$线性无关，从而防止这样的“合并”。

满足这些要求的一个表达式是 $A_4te^{rt}$。因为变量 $t$ 在该式中以积的形式存在，所以它显然与 $A_3e^{rt}$线性无关。这样，它使我们

可以引入另一个常数项 $A_4$。但 $A_4te^{rt}$ 是(15.4)的合适的解吗?如果我们试将其作为(15.4)的解,则由积的微分法则,可求出其一阶及二阶导数为

$$y'(t) = (rt + 1)A_4e^{rt} \quad \text{和} \quad y''(t) = (r^2t + 2r)A_4e^{rt}$$

将 $y$, $y'$, $y''$的表达式代入(15.4)的左边,得到表达式

$$[(r^2t + 2r) + a_1(rt + 1) + a_2t]A_4e^{rt}$$

由于在这里,我们有 $a_1^2 = 4a_2$, $r = -a_1/2$,所以,最后一个表达式恒为零,因而总与(15.4)的右边相等;这证明 $A_4te^{rt}$ 确实可作为(15.4)的解。

因此,二重根情况下的余函数可以写成

$$y_c = A_3e^{rt} + A_4te^{rt} \tag{15.9}$$

**例 5** 解微分方程

509

$$y''(t) + 6y'(t) + 9y = 27$$

这里,系数为 $a_1 = 6$, $a_2 = 9$;因 $a_1^2 = 4a_2$,所以其根为重根。按照公式(15.5),我们有 $r = -a_1/2 = -3$。因此,与(15.9)的结果一致,余函数可以写成

$$y_c = A_3e^{-3t} + A_4te^{-3t}$$

现在,给定函数的通解便容易求出了。试求特别积分的常数解,得到 $y_p = 3$。由此得完备方程的通解为

$$y(t) = y_c + y_p = A_3e^{-3t} + A_4te^{-3t} + 3$$

两个任意常数的值同样可以根据两个初始条件来确定。假设初始条件为 $y(0) = 5$, $y'(0) = -5$。令上述通解中的 $t = 0$,我们应求得 $y(0) = 5$,即

$$y(0) = A_3 + 3 = 5$$

由此得 $A_3=2$，其次，对通解微分，再令 $t=0$ 及 $A_3=2$，必有 $y'(0)=-5$。即

$$y'(t)=-3A_3e^{-3t}-3A_4te^{-3t}+A_4e^{-3t}$$

和 $$y'(0)=-6+A_4=-5$$

由此得 $A_4=1$。因此，我们最终可将给定方程的定解写成

$$y(t)=2e^{-3t}+te^{-3t}+3$$

**第三种情况（复根）** 关于系数 $a_1$ 和 $a_2$ 的相对大小，还存在第三种可能性，即 $a_1^2<4a_2$。当这种可能性万一发生时，公式（15.5）中将包含负数的平方根，在适当引入虚数和复数的概念以前，我们还不能处理这样的问题。由于这个原因，我们将满足于这种分类，并将其留待下节详细讨论。

上述三种情况在图 15.1 中以三条曲线加以说明，每条曲线都表示二次函数 $f(r)=r^2+a_1r+a_2$ 的不同形式。正如我们已了解到的那样，当令此函数等于零时，结果是二次方程 $f(r)=0$，解此方程仅是“求二次函数的零值”。从图形上看，这意味着方程的根可在横轴上 $f(r)=0$ 处求得。 510

图 15.1 中的曲线最下面那条曲线与横轴相交两次，因而我们可以求得两个根 $r_1$ 和 $r_2$，二者均满足二次方程 $f(r)=0$，且均为实数值。因此这条曲线说明了第一种情况。下面转向中间那条曲线，我们注意到它与横轴仅相交于一点 $r_3$。$r_3$ 是唯一的满足方程 $f(r)=0$ 的唯一的值，因此，中间这条曲线说明了第二种情况。最后，我们注意到最上面那条曲线与横轴不相交，不存在满足方程 $f(r)=0$ 的实数根。在此情况下，尽管不存在实数根，但仍存在两个复数，满足这个方程。我们下一节将证明这一点。

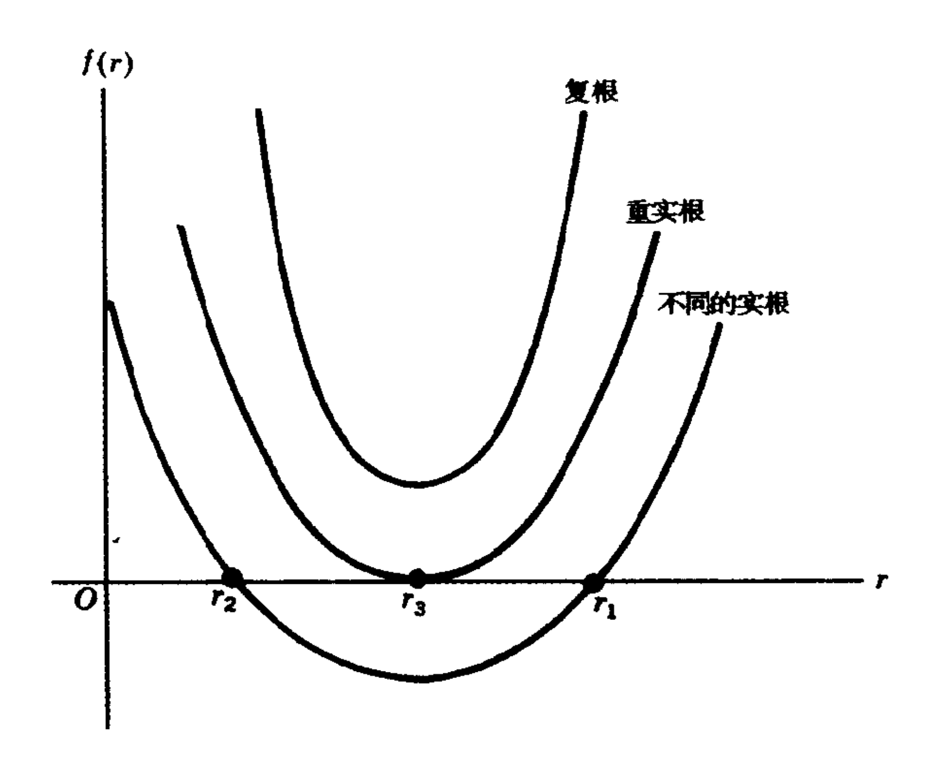



图 15.1

## 均衡的动态稳定性

对于第一、二种情况,均衡的动态稳定性仍取决于特征根的代数符号。

对于第一种情况,余函数(15.7)由两个指数式 $A_1e^{r_1t}$ 和 $A_2e^{r_2t}$构成。系数 $A_1$ 和 $A_2$ 为任意常数,其值取决于问题的初始条件。因此,无论初始条件如何,我们可以确定,当且仅当两个根 $r_1$ 和 $r_2$ 均为负,方可达到动态稳定均衡(当 $t\to\infty$时,$y_c\to0$)。我们强调"两个"一词,是因为动态稳定性的条件甚至不允许其中的一个根为正或者为零。比如,如果 $r_1=2$,而 $r_2=-5$,初看起来,第二个根有较大的绝对值,似乎可以超过第一个。但实际上,是这个正根最终将起支配作用,因为当 $t$ 增大时,$e^{2t}$将会增大,而 $e^{-5t}$

则会逐渐变小。

对于第二种重根的情况，余函数(15.9)不仅包含我们熟悉的 $e^{rt}$ 指数式，而且包含乘积式 $te^{rt}$，不管初始条件如何，$e^{rt}$ 趋近于零的充要条件是 $r<0$。但是什么条件能保证 $te^{rt}$ 为零呢？实际上，$te^{rt}$(或更一般地，$t^k e^{rt}$)与 $e^{rt}(r\neq 0)$ 拥有相同类型的一般时间路径。因此，条件 $r<0$ 确实是当 $t\to\infty$ 时，整个完备方程趋近于零，产生瞬时动态稳定均衡的充要条件。 511

## 练习 15.1

---

1 求每个方程的特别积分：

(a) $y''(t)-2y'(t)+5y=2$

(b) $y''(t)+y'(t)=7$

(c) $y''(t)+3y=9$

(d) $y''(t)+2y'(t)-y=-4$

(e) $y''(t)=12$

2 求每个方程的余函数：

(a) $y''(t)+3y'(t)-4y=12$

(b) $y''(t)+6y'(t)+5y=10$

(c) $y''(t)-2y'(t)+y=3$

(d) $y''(t)+8y'(t)+16y=0$

3 求上题每个微分方程的通解，然后运用初始条件 $y(0)=4, y'(0)=2$，求其定解。

4 上题所求得的瞬时均衡是动态稳定的吗？

5 验证例 5 的定解确实：(a) 满足两个初始条件；(b) 具有与给定微分方程一致的一阶和二阶导数。

6 证明当 $t\to\infty$ 时，若 $r<0$，则 $te^{rt}$ 的极限为零；但若 $r\geq 0$，则其极限为无穷大。

---

# 15.2 复数和三角函数

当二阶线性微分方程 $y''(t)+a_1y'(t)+a_2y=b$ 的系数使得 $a_1^2<4a_2$ 时，特征根公式(15.5)要求取负数的平方根。因为任意正实数或负实数的平方均为正，而零的平方仍为零，所以只有非负实数才会产生实数值平方根。因此，如果我们仅将注意力集中于实数体系，正如我们迄今一直所做的那样，那么就没有特征根适用于这种情况(第三种情况)。这趋使我们去考察实数体系以外的数。

## 虚数与复数

就概念而言，我们可以定义一个数 $i\equiv\sqrt{-1}$，其平方等于 $-1$。由于 $i$ 是负数的平方根，所以它显然不是一个实数。因而，我们称其为虚数。根据这个定义，我们还可以写出许多其它虚数，如 $\sqrt{-9}=\sqrt{9}\sqrt{-1}=3i$ 及 $\sqrt{-2}=\sqrt{2}i$ 等。

进一步扩展虚数概念的应用，我们还可以构造另一种数——512 一种既包含实数部分，也包含虚数部分的数，比如 $(8+i)$ 和 $(3+5i)$。我们将这种数称作复数，一般可以用 $(h+vi)$ 的形式来表示，其中 $h$ 和 $v$ 为两个实数。[①] 当然，当 $v=0$ 时，复数就简化为实数，而若 $h=0$，则复数就化为虚数。因此，所有实数的集合(称其为 **R**)构成了所有复数的集合(称为 **C**)的一个子集。类似地，

---

① 在一般复数符号中，我们使用符号 $h$(表示横向，horizontal)和 $v$(表示纵向，vertical)，因为我们将在二维图形的横轴和纵轴上分别描绘 $h$ 和 $v$ 的值。

所有虚数的集合(称其为 **I**)也构成了 **C** 的一个子集。即 $\mathbf{R}\subset\mathbf{C}$ 和 $\mathbf{I}\subset\mathbf{C}$。而且因为“实”与“虚”这两个术语相互排斥,所以集合 **R** 和 **I** 必不相交,即 $\mathbf{R}\cap\mathbf{I}=\varnothing$。

复数$(h+vi)$在几何上可通过阿拉贡图来表示,如图 15.2 所示,把 $h$ 水平地绘在实轴上,$v$ 垂直地绘在虚轴上,这样,复数$(h+vi)$可以点$(h,v)$来表示,在图中我们还将此点标为 $C$。当然,$h$ 和 $v$ 值具有代数符号;若 $h<0$,点 $C$ 位于原点的左边;类似地,负的 $v$ 意味着点 $C$ 位于横轴下方。

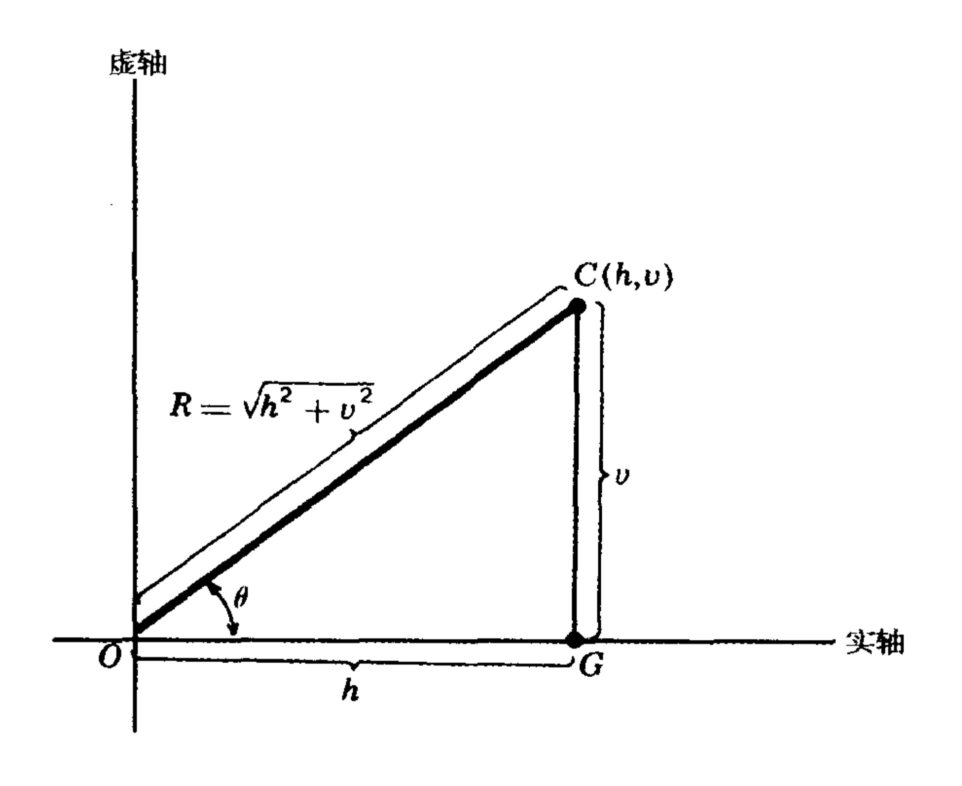



图 15.2

给定 $h$ 和 $v$ 值,运用毕达哥拉斯定理,[①]我们也可以计算出直线 $OC$ 的长度,即直角三角形斜边的平方等于另外两边平方之和。以 $R$(表示向量径)表示 $OC$ 的长度,我们有

① 即勾股定理。——译者

$$R^2 = h^2 + v^2 \text{ 和 } R = \sqrt{h^2 + v^2} \qquad (15.10)$$

其中平方根总取正值。$R$ 的值有时也称作复数$(h + vi)$的**绝对值**或**模**。(注意,改变 $h$ 和 $v$ 的符号对复数的绝对值 $R$ 无影响。)同 513 $h$ 和 $v$ 一样,$R$ 也是实数值,但与其不同的是,$R$ 总为正。我们将发现,数 $R$ 在下面的讨论中具有重要意义。

## 复 根

在此过程中,让我们回到公式(15.5),考察复特征根的情况。当二阶微分方程的系数使得 $a_1^2 < 4a_2$ 时,平方根表达式(15.5)可以写成

$$\sqrt{a_1^2 - 4a_2} = \sqrt{4a_2 - a_1^2}\sqrt{-1} = \sqrt{4a_2 - a_1^2}\, i$$

因此,若采用简写符号

$$h = \frac{-a_1}{2} \quad \text{和} \quad v = \frac{\sqrt{4a_2 - a_1^2}}{2}$$

则两个根可用一对**共轭复数**来表示:

$$r_1,\ r_2 = h \pm vi$$

这两个复根之所以称作"共轭",是因为二者总是同时出现,一个是 $h$ 与 $vi$ 的**和**,另一个为 $h$ 与 $vi$ 的**差**。注意,它们具有相同的绝对值 $R$。

**例 1** 求特征方程 $r^2 + r + 4 = 0$ 的根。应用熟悉的公式,得到

$$r_1,\ r_2 = \frac{-1 \pm \sqrt{-15}}{2} = \frac{-1 \pm \sqrt{15}\sqrt{-1}}{2} = \frac{-1}{2} \pm \frac{\sqrt{15}}{2} i$$

它们构成了一对共轭复数。

同以前一样,可以使用(15.6)来验证我们的计算结果。若计

算正确,则我们应有 $r_1+r_2=-a_1(=-1)$ 和 $r_1r_2=a_2(=4)$。因为确实可求得

$$r_1+r_2=\left(\frac{-1}{2}+\frac{\sqrt{15}\,i}{2}\right)+\left(\frac{-1}{2}-\frac{\sqrt{15}\,i}{2}\right)$$

$$=\frac{-1}{2}+\frac{-1}{2}=-1$$

和

$$r_1r_2=\left(\frac{-1}{2}+\frac{\sqrt{15}\,i}{2}\right)\left(\frac{-1}{2}-\frac{\sqrt{15}\,i}{2}\right)$$

$$=\left(\frac{-1}{2}\right)^2-\left(\frac{\sqrt{15}\,i}{2}\right)^2=\frac{1}{4}-\frac{-15}{4}=4$$

所以,我们的计算结果是正确的。

即使在复根的情况下(第三种情况),我们仍可按(15.7)来表示微分方程的余函数。即

$$y_c=A_1e^{(h+vi)t}+A_2e^{(h-vi)t}=e^{ht}(A_1e^{vit}+A_2e^{-vit}) \tag{15.11}$$

514

但这里已引进了一个新的特点:数 $i$ 出现在括号中两个指数的表达式中。我们如何解释这个虚指数函数呢?

为便于解释,先将这些指数式转化为等价的三角函数是有益的。正如我们将要看到的那样,三角函数具有变量周期波动的特征。因此,作为可转化为三角函数的余函数(15.11),可以预期它将产生周期性的时间路径。

## 三角函数

如图 15.3 所示,考察以原点为圆心,以 $R$ 为半径的圆。假设半径像时针一样逆时针由 $OA$ 开始旋转,逐渐旋转到 $OP$,再依次旋转到位置 $OB$, $OC$ 及 $OD$,最后再旋转至 $OA$。以后,圆又可以

重复旋转。

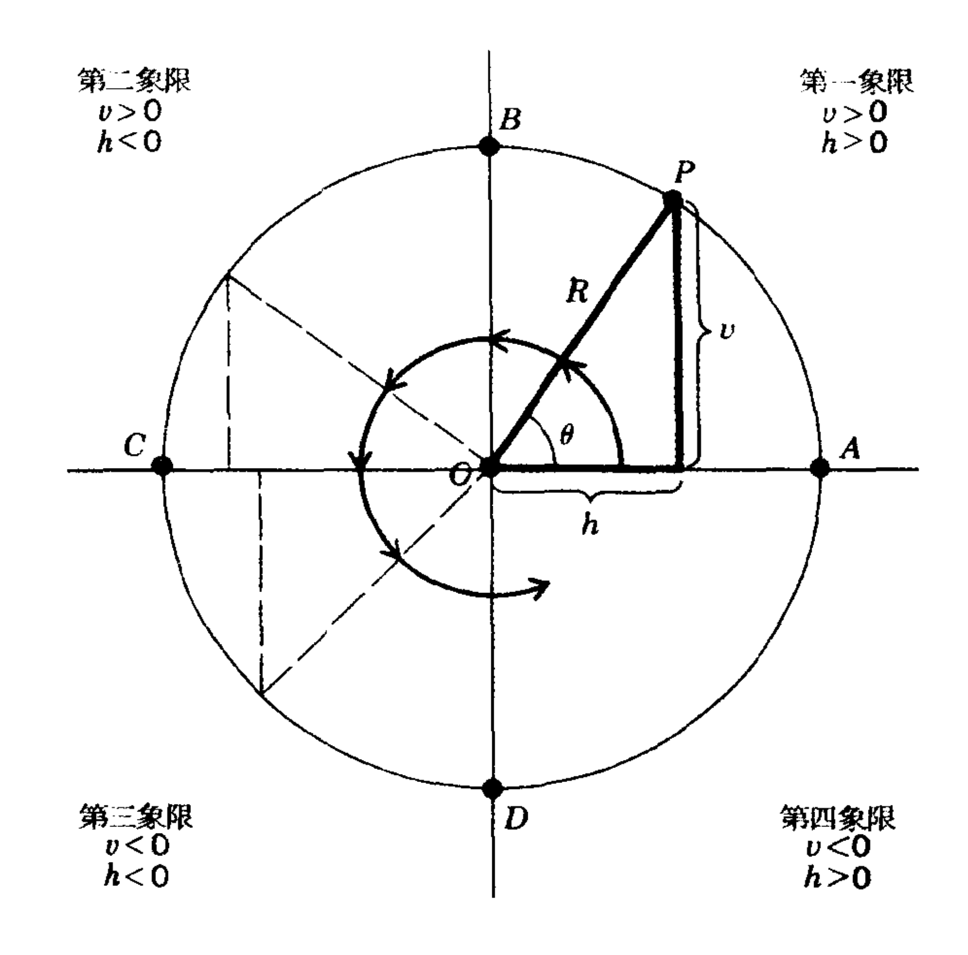



图 15.3

当时针旋转至某一特定位置,比如 $OP$ 时,与线段 $OA$ 形成定角 $\theta$,时针的端点($P$)确定垂直距离 $v$ 和水平距离 $h$。在旋转过程中,当角度 $\theta$ 变化时,尽管 $R$ 不变,但 $v$ 和 $h$ 会发生变化。这样,比率 $v/R$ 和 $h/R$ 必随 $\theta$ 变化而变化,即此两个比率均是角 $\theta$ 的函数。具体地说,$v/R$ 和 $h/R$ 分别称作 $\theta$ 的**正弦**(函数)和**余弦**(函数)

$$\sin\theta = \frac{v}{R} \tag{15.12}$$

$$\cos\theta = \frac{h}{R} \tag{15.13}$$

由于这些函数与圆相联系，所以它们被称作圆函数。又因为它们还与三角形有关，所以也被称为三角函数。它们还有另一个别致的名称——类正余弦函数。正弦函数和余弦函数并非仅有的三角函数，另一个常见的三角函数是正切函数，定义为

$$\tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{v}{h} \quad (h \neq 0)$$

但我们这里主要关心的是正弦函数和余弦函数。

三角函数的自变量是角 $\theta$，因而这里的映射是从角到两个距离比率的映射。通常，角以度数（如 30°，45°，90°等）来度量；但在分析工作中，用弧度来度量更方便一些。弧度度量的优点源于下述事实：当 $\theta$ 以弧度度量时，三角函数的导数会产生更简洁的表达式——这非常类似于以 $e$ 为底的指数函数和对数函数。但一弧度等于多少呢？为解释这一问题，我们再回到图 15.3，在图中我们已绘出了点 $P$，使得弧 $AP$ 的长度恰好等于半径 $R$ 的长度。以这样的弧长 $R$ 所定义的角 $\theta$ 的大小，称作一弧度（简写为 rad）。因为圆周长等于 $2\pi R$（其中 $\pi = 3.14159\cdots\cdots$），所以一个完整的圆必定等于 $2\pi$ 弧度。但按度数来度量，一个完整的圆为 360°。因此，令 360°等于 $2\pi$ 弧度，我们可以得到下述换算表： 515

| 度数 | 360 | 270 | 180 | 90 | 45 | 0 |
|---|---|---|---|---|---|---|
| 弧度 | $2\pi$ | $\frac{3\pi}{2}$ | $\pi$ | $\frac{\pi}{2}$ | $\frac{\pi}{4}$ | 0 |

## 正、余弦函数的性质

给定 $R$ 的长度，$\sin\theta$ 的值取决于 $v$ 值随角 $\theta$ 变化的方式。在起点 $OA$，$v = 0$。当时针逆时针方向旋转时，$v$ 值递增，并取正值，当时针与 $OB$ 重合，即当 $\theta = \pi/2$（$= 90°$）时，达到最大值 $v = R$。 516

继续旋转，$v$ 值开始递减，直至时针处于 $OC$，即 $\theta=\pi$（$=180^\circ$）时，$v$ 值变为零。当时针进入第三象限，$v$ 开始取负值，在 $OD$ 处，$v=-R$。在第四象限，$v$ 仍为负，但当时针返回到 $OA$ 时，它从负值（$-R$）增至 $v=0$，即当 $\theta=2\pi$（$=360^\circ$）时，$v=0$。再继续旋转，结果会重复。

当把 $v$ 值代入(15.12)时，可得到在表 15.1 中"$\sin\theta$"一行表示的结果。但为更全面地描述正弦函数，请参看图 15.4($a$)中的图形。在该图中依据 $\theta$ 值（以弧度表示）绘出 $\sin\theta$ 的值。

**表 15.1**

| $\theta$ | 0 | $\frac{1}{2}\pi$ | $\pi$ | $\frac{3}{2}\pi$ | $2\pi$ |
|---|---|---|---|---|---|
| $\sin\theta$ | 0 | 1 | 0 | $-1$ | 0 |
| $\cos\theta$ | 1 | 0 | $-1$ | 0 | 1 |

与此相反，$\cos\theta$ 的值取决于随 $\theta$ 而变化的 $h$ 值。在起始位置 $OA$，有 $h=R$。然后，$h$ 逐渐减小，当 $\theta=\pi/2$（位置 $OB$）时，$h=0$。在第二象限，$h$ 为负，当 $\theta=\pi$（位置 $OC$）时，$h=-R$。在第三象限，$h$ 由 $-R$ 增至零，当 $\theta=3\pi/2$（位置 $OD$）时，$h=0$。在第四象限，$h$ 又变为正值，当时针返回到位置 $OA$ 时（$\theta=2\pi$），我们又有 $h=R$。然后圆又重新旋转。

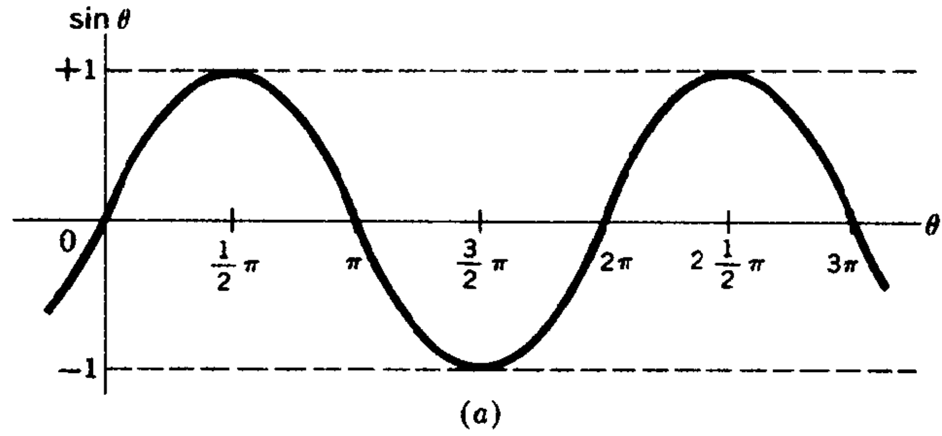
将上面那些 $h$ 值代入(15.13)，得到表 15.1 中最下边一行的结果，但图 15.4$b$ 余弦函数给出了更完全的描述。

函数 $\sin\theta$ 和 $\cos\theta$ 具有相同的定义域，即所有实数（$\theta$ 的弧度度量）的集合。在这方面，也许我们应该指出，负角不过意味着时针顺时针旋转。比如，在图 15.3 中，从 $OA$ 到 $OD$ 的顺时针旋转，产生 $-\pi/2$ 弧度（$=-90^\circ$）。这两个函数也有共同的值域，即闭区

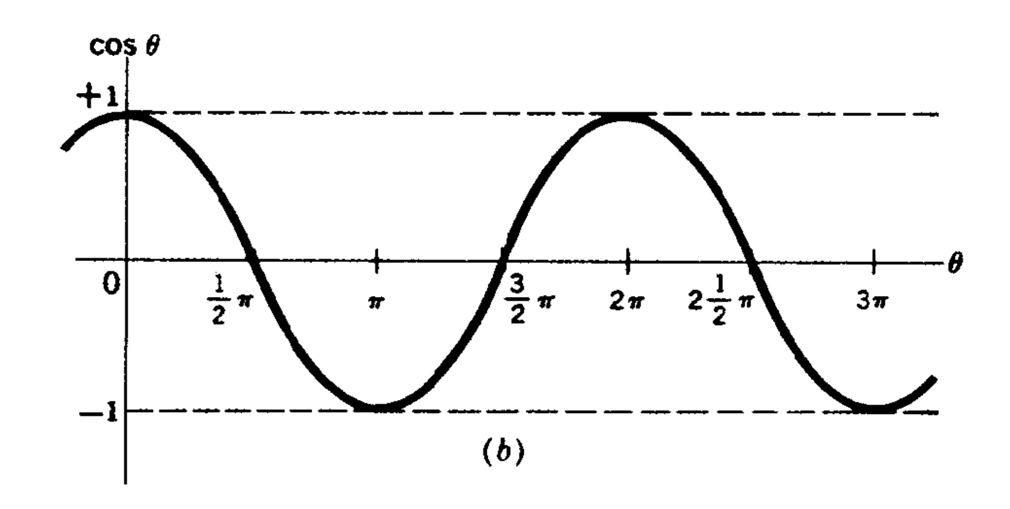
间[ − 1, 1]。因此,图 15.4 中 $\sin\theta$ 和 $\cos\theta$ 的图形限定在一定的水平带内。

517






(a)




(b)

图 15.4

正弦函数和余弦函数的主要特征是它们均具有**周期性**,角 $\theta$ 旋转 $2\pi$ 弧度,函数值重复一次。因此,我们可以说,每个函数以 $2\pi$ 为周期。基于这种周期性特征,下列方程成立(对任意整数 $n$):

$$\sin(\theta + 2n\pi) = \sin\theta \quad \cos(\theta + 2n\pi) = \cos\theta$$

即任意角 $\theta$ 加上或减去 $2\pi$ 的任意整数倍,对 $\sin\theta$ 和 $\cos\theta$ 的值均无影响。

正、余弦函数的图形表明,函数在每一波动周期都有固定的波

幅,即±1。所以我们有时也以函数的波幅为1来描述这种特征。由于 $\sin\theta$ 和 $\cos\theta$ 的波幅相等,周期相同,所以,我们看到,如果将 $\cos\theta$ 曲线右移 $\frac{\pi}{2}$,则它将与 $\sin\theta$ 曲线恰好重合。因此,我们说这两条曲线的差别仅在于周相不同,即在每个周期中峰值的位置不同。这个事实用符号可用下列方程来描述:

$$\cos\theta = \sin\left(\theta + \frac{\pi}{2}\right)$$

正弦和余弦函数满足某些恒等式。其中比较常见的有

$$\sin(-\theta) \equiv -\sin\theta$$

$$\cos(-\theta) \equiv \cos\theta \tag{15.14}$$

$$\sin^2\theta + \cos^2\theta \equiv 1 \ [\text{where } \sin^2\theta \equiv (\sin\theta)^2, \text{等。}]$$

$$\sin(\theta_1 \pm \theta_2) \equiv \sin\theta_1\cos\theta_2 \pm \cos\theta_1\sin\theta_2 \tag{15.15}$$

$$\cos(\theta_1 \pm \theta_2) \equiv \cos\theta_1\cos\theta_2 \mp \sin\theta_1\sin\theta_2 \tag{15.16}$$

(15.14)中的两个恒等式用于强调这一事实:余弦函数是关于纵轴对称的(即 $\theta$ 与 $-\theta$ 得到同样的余弦值,但正弦函数对纵轴是非对称的。(15.15)中的恒等式表明,对任意 $\theta$ 值,其正弦和余弦函数的平方和恒等于1。(15.16)中的一组恒等式给出了两个角 $\theta_1$ 和 $\theta_2$ 的和与差的正弦与余弦。这些恒等式在后面的讨论中是极有益的。

518

最后,谈一下导数。正、余弦函数均是连续、平滑和可导的。当 $\Delta\theta\to 0$,分别取微商 $\Delta(\sin\theta)/\Delta\theta$,$\Delta(\cos\theta)/\Delta\theta$ 的极限,便可得到导数 $d(\sin\theta)/d\theta$ 和 $d(\cos\theta)/d\theta$。这里仅叙述结果,不加以证明:

$$\frac{d}{d\theta}\sin\theta = \cos\theta \tag{15.17}$$

$$\frac{d}{d\theta}\cos\theta = -\sin\theta \qquad (15.18)$$

但应强调的是，仅当 $\theta$ 以弧度来度量时，导数公式方成立。若以度数来度量，如(15.17)会变成 $d(\sin\theta)/d\theta=(\pi/180)\cos\theta$。正是为了去掉因子$(\pi/180)$，所以在分析工作中宁愿选用弧度，而不用度数作为度量单位。

**例 2** 求 $\sin\theta$ 曲线在 $\theta=\pi/2$ 的斜率。正弦曲线的斜率由其导数$(=\cos\theta)$给出。因此，在 $\theta=\pi/2$，斜率应等于 $\cos(\pi/2)=0$。读者可由图 15.4 来验证这一结果。

**例 3** 求 $\sin\theta$ 的二阶导数。由(15.17)，我们知道，$\sin\theta$ 的一阶导数为 $\cos\theta$，因此，所求一阶导数为

$$\frac{d^2}{d\theta^2}\sin\theta = \frac{d}{d\theta}\cos\theta = -\sin\theta$$

## 欧拉关系

在 9.5 节中，我们已证明任意具有有限、连续直至所需阶导数的函数可以展开成一个多项式函数。而且，如果所产生的泰勒级数(在任意点 $x_0$ 展开)或麦克劳林级数(在 $x_0=0$ 处展开)的余项，在项数 $n$ 变得无穷大时恰好为零，则该多项式可以写成无穷级数。现在我们先展开正弦和余弦函数，然后再设法证明我们在(15.11)中所遇到的虚指数式可以变换成具有相同展开式的三角函数。

对于正弦函数，写成 $\phi(\theta)=\sin\theta$。由此得 $\phi(0)=\sin 0=0$。519
连续求导，得到

$$\left.\begin{array}{l}\phi'(\theta)=\cos\theta \\ \phi''(\theta)=-\sin\theta \\ \phi'''(\theta)=-\cos\theta \\ \phi^{(4)}(\theta)=\sin\theta \\ \phi^{(5)}(\theta)=\cos\theta \\ \vdots \qquad \vdots\end{array}\right\} \Rightarrow \left\{\begin{array}{l}\phi'(0)=\cos 0=1 \\ \phi''(0)=-\sin 0=0 \\ \phi'''(0)=-\cos 0=-1 \\ \phi^{(4)}(0)=\sin 0=0 \\ \phi^{(5)}(0)=\cos 0=1 \\ \vdots \qquad \vdots\end{array}\right.$$

当将其代入(9.14),其中 $\theta$ 现在代替了 $x$,得到如下带余项的麦克劳林展开式:

$$\sin\theta=0+\theta+0-\frac{\theta^3}{3!}+0+\frac{\theta^5}{5!}+\cdots+\frac{\phi^{(n+1)}(p)}{(n+1)!}\theta^{n+1}$$

现在,最后一项(余项)中的 $\phi^{(n+1)}(p)$,表示 $(n+1)$ 阶导数在 $\theta=p$ 处计算的值,仅能取 $\pm\cos p$ 或 $\pm\sin p$ 的形式,因此,仅能在区间 $[-1,1]$ 中取值,而与 $n$ 的大小无关。另一方面,$(n+1)!$ 当 $n\to\infty$ 时会迅速增加——事实上,当 $n$ 增加时,$(n+1)!$ 要比 $\theta^{n+1}$ 增加得更快。因此,当 $n\to\infty$ 时,余项将趋于零,因此,我们可以将麦克劳林级数表示成一个无穷级数:

$$\sin\theta=\theta-\frac{\theta^3}{3!}+\frac{\theta^5}{5!}-\frac{\theta^7}{7!}+\cdots \tag{15.19}$$

类似地,若记 $\psi(\theta)=\cos\theta$,则 $\psi(0)=\cos 0=1$,其各阶导数依次为

$$\left.\begin{array}{l}\psi'(\theta)=-\sin\theta \\ \psi''(\theta)=-\cos\theta \\ \psi'''(\theta)=\sin\theta \\ \psi^{(4)}(\theta)=\cos\theta \\ \psi^{(5)}(\theta)=-\sin\theta \\ \vdots \qquad \vdots\end{array}\right\} \Rightarrow \left\{\begin{array}{l}\psi'(0)=-\sin 0=0 \\ \psi''(0)=-\cos 0=-1 \\ \psi'''(0)=\sin 0=0 \\ \psi^{(4)}(0)=\cos 0=1 \\ \psi^{(5)}(0)=-\sin 0=0 \\ \vdots \qquad \vdots\end{array}\right.$$

基于这些导数,可将 $\cos\theta$ 展开如下:

$$\cos\theta = 1 + 0 - \frac{\theta^2}{2!} + 0 + \frac{\theta^4}{4!} + \cdots + \frac{\psi^{(n+1)}(p)}{(n+1)!}\theta^{n+1}$$

因为当 $n \to \infty$ 时,余项仍趋近于零,则余弦函数也可以表示成如下 520 无穷级数:

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots \tag{15.20}$$

读者一定注意到,有了(15.19)和(15.20)式,对所有可能的 $\theta$ 值(弧度),我们现在可以构造一个正弦值和余弦值的表。但我们眼前关心的是求出虚指数式与三角函数的关系。为此,我们现在展开两个指数式 $e^{i\theta}$ 和 $e^{-i\theta}$。读者将会认识到它们是式 $e^x$ 的特殊形式。在(10.6)中已给出了 $e^x$ 的展开式

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \cdots$$

因此,令 $x = i\theta$,立即得到

$$\begin{aligned} e^{i\theta} &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \frac{(i\theta)^5}{5!} + \cdots \\ &= 1 + i\theta - \frac{\theta^2}{2!} - \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{i\theta^5}{5!} - \cdots \\ &= (1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots) + i(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots) \end{aligned}$$

类似地,令 $x = -i\theta$,可得到如下结果

$$\begin{aligned} e^{-i\theta} &= 1 - i\theta + \frac{(-i\theta)^2}{2!} + \frac{(-i\theta)^3}{3!} + \frac{(-i\theta)^4}{4!} + \frac{(-i\theta)^5}{5!} + \cdots \\ &= 1 - i\theta - \frac{\theta^2}{2!} + \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} - \frac{i\theta^5}{5!} - \cdots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots\right) - i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots\right) \end{aligned}$$

将(15.19)和(15.20)代入上述两个结果,便可确立如下两个恒等式(称作欧拉关系):

$$e^{i\theta} \equiv \cos\theta + i\sin\theta \tag{15.21}$$

$$e^{-i\theta} \equiv \cos\theta - i\sin\theta \tag{15.21'}$$

这两个公式使我们可以将任意虚指数函数转换成等价的正弦和余弦函数的线性组合,反之亦然。

**例4** 求 $e^{i\pi}$ 的值。我们首先将此式转换成三角函数式。令(15.21)中的 $\theta=\pi$,可求得 $e^{i\pi}=\cos\pi+i\sin\pi$。因 $\cos\pi=-1$,$\sin\pi=0$,由此得 $e^{i\pi}=-1$。

**例5** 证明 $e^{-i\pi/2}=-i$。令(15.21′)中的 $\theta=\pi/2$,我们有:

$$e^{-i\pi/2}=\cos\frac{\pi}{2}-i\sin\frac{\pi}{2}=0-i(1)=-i$$

## 复数的另一种表示方式

521 迄今为止,我们一直把一对共轭复数以一般形式$(h\pm vi)$来表示。因为在一个阿拉贡图形的笛卡尔坐标系中,$h$ 和 $v$ 表示横坐标与纵坐标,所以,表达式$(h\pm vi)$表示一对共轭复数的**笛卡尔形式**。作为讨论三角函数和欧拉关系的副产品,我们现在可以另外两种方式表示$(h\pm vi)$。

参考图15.2,一旦 $h$ 和 $v$ 的值确定,则角 $\theta$ 和 $R$ 的值也就确定了。因为已知的 $\theta$ 和 $R$ 可共同在阿拉贡图中确定一个点,所以,我们可使用 $\theta$ 和 $R$ 确定一对特定的共轭复数。将正弦函数(15.12)和余弦函数(15.13)的定义重写成;

$$v=R\sin\theta \text{ 和 } h=R\cos\theta \tag{15.22}$$

共轭复数$(h\pm vi)$可以变换如下:

$$h \pm vi = R\cos\theta \pm Ri\sin\theta = R(\cos\theta \pm i\sin\theta)$$

在变换时，我们实际上已由复数的笛卡尔坐标（$h$ 和 $v$）转换为所谓的极坐标（$R$ 和 $\theta$）。相应地，上面方程右边的表达式给出了共轭复数的极形式。

进而，由欧拉关系，极形式也可以写成如下指数形式：$R(\cos\theta \pm i\sin\theta) = Re^{\pm i\theta}$。因此，我们有共轭复数的三种可供选择的表达方式：

$$h \pm vi = R(\cos\theta \pm i\sin\theta) = Re^{\pm i\theta} \tag{15.23}$$

若已知 $R$ 和 $\theta$ 的值，我们可利用(15.22)中的两个方程，直接变换为 $h$ 和 $v$。那么，逆变换又如何呢？根据已知的 $h$ 和 $v$ 值，求相应的 $R$ 值并无困难，它等于 $\sqrt{h^2+v^2}$。但求 $\theta$ 则有些困难；所求的 $\theta$ 值（弧度）满足两个条件 $\cos\theta = h/R$ 和 $\sin\theta = v/R$；但对给定的 $h$ 和 $v$ 值，$\theta$ 不是唯一的（为什么?）。好在这个问题并不严重，把我们的注意力集中于定义域区间 $(0, 2\pi)$，不确定性便可以解决。

**例 6** 求复数 $5e^{3i\pi/2}$ 的笛卡尔形式。这里我们有 $R=5, \theta=3\pi/2$，因此，根据(15.22)和表 15.1，

$$h = 5\cos\frac{3\pi}{2} = 0 \quad 和 \quad v = 5\sin\frac{3\pi}{2} = -5$$

因此，笛卡尔形式为 $h + vi = -5i$。

**例 7** 求 $(1+\sqrt{3}i)$ 的极形式和指数形式。在本例中，我们有 $h=1$，$v=\sqrt{3}$，因此 $R=\sqrt{1+3}=2$。这次表 15.1 中的值对确定 $\theta$ 值无帮助，但列出了另外一些 $\sin\theta$ 和 $\cos\theta$ 值的表 15.2 是有帮助的。具体地说，我们在求使得 $\cos\theta = h/R = 1/2$ 和 $\sin\theta = v/R = \sqrt{3}/2$ 的 $\theta$ 值。值 $\theta = \pi/3$ 满足这一要求。因此，根据(15.23)，所 522

求变换为

$$1+\sqrt{3}i=2(\cos\frac{\pi}{3}+i\sin\frac{\pi}{3})=2e^{i\pi/3}$$

在结束这个话题以前，我们还需注意一下结果(15.23)的一个重要推广。假设我们有一个复数的 $n$ 次幂，比如 $(h+vi)^n$，那么，我们如何写出其极形式和指数形式？指数形式是很容易推导的。因为 $h+vi=Re^{i\theta}$，由此得

$$(h+vi)^n=(Re^{i\theta})^n=R^ne^{in\theta}$$

类似地，我们可以写出

$$(h-vi)^n=(Re^{-i\theta})^n=R^ne^{-in\theta}$$

注意，幂数 $n$ 已发生两种变化：(1) $R$ 现在变成了 $R^n$；(2) $\theta$ 现在变成了 $n\theta$。将这两种变化加入(15.23)的极形式中时，我们求得

$$(h\pm vi)^n=R^n(\cos n\theta\pm i\sin n\theta) \qquad (15.23')$$

即

$$[R(\cos\theta\pm i\sin\theta)]^n=R^n(\cos n\theta\pm i\sin n\theta)$$

此即棣莫弗定理(*De Moivre's Theorem*)。该定理表明，复数自乘 $n$ 次幂，只需修正极坐标如下：$R$ 自乘 $n$ 次幂，并以 $n$ 乘 $\theta$。

**表 15.2**

| $\theta$ | $\frac{\pi}{6}$ | $\frac{\pi}{4}$ | $\frac{\pi}{3}$ | $\frac{3\pi}{4}$ |
|---|---|---|---|---|
| $\sin\theta$ | $\frac{1}{2}$ | $\frac{1}{\sqrt{2}}\left(=\frac{\sqrt{2}}{2}\right)$ | $\frac{\sqrt{3}}{2}$ | $\frac{1}{\sqrt{2}}\left(=\frac{\sqrt{2}}{2}\right)$ |
| $\cos\theta$ | $\frac{\sqrt{3}}{2}$ | $\frac{1}{\sqrt{2}}\left(=\frac{\sqrt{2}}{2}\right)$ | $\frac{1}{2}$ | $\frac{-1}{\sqrt{2}}\left(=\frac{-\sqrt{2}}{2}\right)$ |

## 练习 15.2

1 求如下二次方程的根

(a) $r^2-3r+9=0$　　(c) $2x^2+x+8=0$

(b) $r^2+2r+17=0$　　(d) $2x^2-x+1=0$

2 (a)一弧度等于多少度?

(b)一度等于多少弧度?

3 参考图 15.3,并运用勾股定理,证明

523

(a) $\sin^2\theta+\cos^2\theta\equiv1$　　(b) $\sin\frac{\pi}{4}=\cos\frac{\pi}{4}=\frac{1}{\sqrt{2}}$

4 运用恒等式(15.14),(15.15)及(15.16),证明:

(a) $\sin2\theta\equiv2\sin\theta\cos\theta$

(b) $\cos2\theta\equiv1-2\sin^2\theta$

(c) $\sin(\theta_1+\theta_2)+\sin(\theta_1-\theta_2)\equiv2\sin\theta_1\cos\theta_2$

(d) $1+\tan^2\theta\equiv\frac{1}{\cos^2\theta}$

(e) $\sin(\frac{\pi}{2}-\theta)\equiv\cos\theta$

(f) $\cos(\frac{\pi}{2}-\theta)\equiv\sin\theta$

5 运用链式求导法则:

(a) 写出$\frac{d}{d\theta}\sin f(\theta)$和$\frac{d}{d\theta}\cos f(\theta)$的导数公式,其中 $f(\theta)$为 $\theta$ 的函数。

(b) 求 $\cos\theta^3$,$\sin(\theta^2+3\theta)$,$\cos e^\theta$ 及 $\sin(1/\theta)$的导数。

6 由欧拉关系,导出:

(a) $e^{-i\pi}=-1$　　(c) $e^{i\pi/4}=\frac{\sqrt{2}}{2}(1+i)$

(b) $e^{i\pi/3}=\frac{1}{2}(1+\sqrt{3}i)$　　(d) $e^{-3i\pi/4}=-\frac{\sqrt{2}}{2}(1+i)$

7 求下列每个复数的笛卡尔形式:

(a) $2\left(\cos\frac{\pi}{6}+i\sin\frac{\pi}{6}\right)$　(b) $4e^{i\pi/3}$　(c) $\sqrt{2}e^{-i\pi/4}$

8 求下列复数的极式和指数形式:

(a) $\frac{3}{2}+\frac{3\sqrt{3}}{2}i$　　(b) $4(\sqrt{3}+i)$

---

## 15.3 复根情况的分析

掌握了复数和三角函数的概念以后，我们现在准备分析在15.1节中提到的复根的情况（第三种情况）。读者回忆一下，根据特征根的性质将其分为三种类型，仅与微分方程的余函数有关。所以，我们将注意力继续集中于简化方程

$$y''(t)+a_1y'(t)+a_2y=0 \qquad [(15.4)\text{的重述}]$$

### 余函数

当系数 $a_1$ 和 $a_2$ 使得 $a_1^2<4a_2$ 时，特征根将是一对共轭复数

$$r_1,\ r_2=h\pm vi$$

其中　　$h=-\frac{1}{2}a_1$　和　$v=\frac{1}{2}\sqrt{4a_2-a_1^2}$

524 则正如我们早已看到的那样，余函数的形式为

$$y_c=e^{ht}(A_1e^{vit}+A_2e^{-vit}) \qquad [(15.11)\text{的重述}]$$

首先，我们将括号中的虚指数表达式变换为等价的三角函数表达式，从而可以将余函数解释为三角函数。利用欧拉关系可完成这一变换。令(15.21)和(15.21′)中的 $\theta=vt$，我们可以求得

$$e^{vit}=\cos vt+i\sin vt \qquad \text{和} \qquad e^{-vit}=\cos vt-i\sin vt$$

由此可知，(15.11)中的余函数可以重写成

$$y_c=e^{ht}[A_1(\cos vt+i\sin vt)+A_2(\cos vt-i\sin vt)]$$

$$= e^{ht}[(A_1 + A_2)\cos vt + (A_1 - A_2)i\sin vt] \tag{15.24}$$

进而,若我们运用简写符号

$$A_5 \equiv A_1 + A_2 \quad 和 \quad A_6 \equiv (A_1 - A_2)i$$

可以将(15.24)简化为①

$$y_c = e^{ht}(A_5\cos vt + A_6\sin vt) \tag{15.24'}$$

其中 $A_5$ 和 $A_6$ 为新的任意常数,它们有待于在后面加以确定。

细心的读者会感觉到,在前面的步骤中以 $vt$ 代替 $\theta$ 有些不自然。变量 $\theta$ 是角度的度量,而 $vt$ 则是 $t$ 单位(在这里是时间)的大小。那么,我们怎能作 $vt = \theta$ 这种代换呢? 这个问题可以根据图 15.5 中的单位圆(半径 $R = 1$ 的圆)来解释。确实,我们一直以 $\theta$ 来表示角,但因为角是以弧度为单位来度量的,所以 $\theta$ 值总是弧 $AB$ 的长度与半径 $R$ 的比率。当 $R = 1$ 时,具体地有

$$\theta \equiv \frac{\text{arc}AB}{R} \equiv \frac{\text{arc}AB}{1} \equiv \text{arc}AB$$

换言之,$\theta$ 不仅是角的弧度度量,而且也是弧 $AB$ 长度的度量。而弧 $AB$ 是一个数而非一个角。如果过去的时间不是像描述时间序列那样标在直线上,而是标在单位圆(逆时针)的圆周上,那么,将时间的流逝视为角 $\theta$ 弧度的增加,还是看成弧 $AB$ 长度的增加,是没有什么差别的。而且,即使 $R \neq 1$,也可以应用同样的原理,只

---

① $A_6$ 的定义中包含虚数 $i$,这并不意味着我们试图"将垃圾盖在地毯下面"。因为 $A_6$ 为一任意常数,所以它既可以取虚数,也可以取实数。根据定义,它不一定必须取虚数。实际上,若 $A_1$ 和 $A_2$ 为一对共轭复数,比如 $m \pm ni$,则 $A_5$ 和 $A_6$ 将均为实数:$A_5 = A_1 + A_2 = (m + ni) + (m - ni) = 2m$,$A_6 = (A_1 + A_2)i = [(m + ni) - (m - ni)]i = (2ni)i = -2n$。

不过在此情况下，$\theta=(\text{弧 } AB)/R$，即角 $\theta$ 和弧 $AB$ 维持固定比例，525 而不是彼此相等。因此，代换 $\theta=vt$ 的确是合理的。

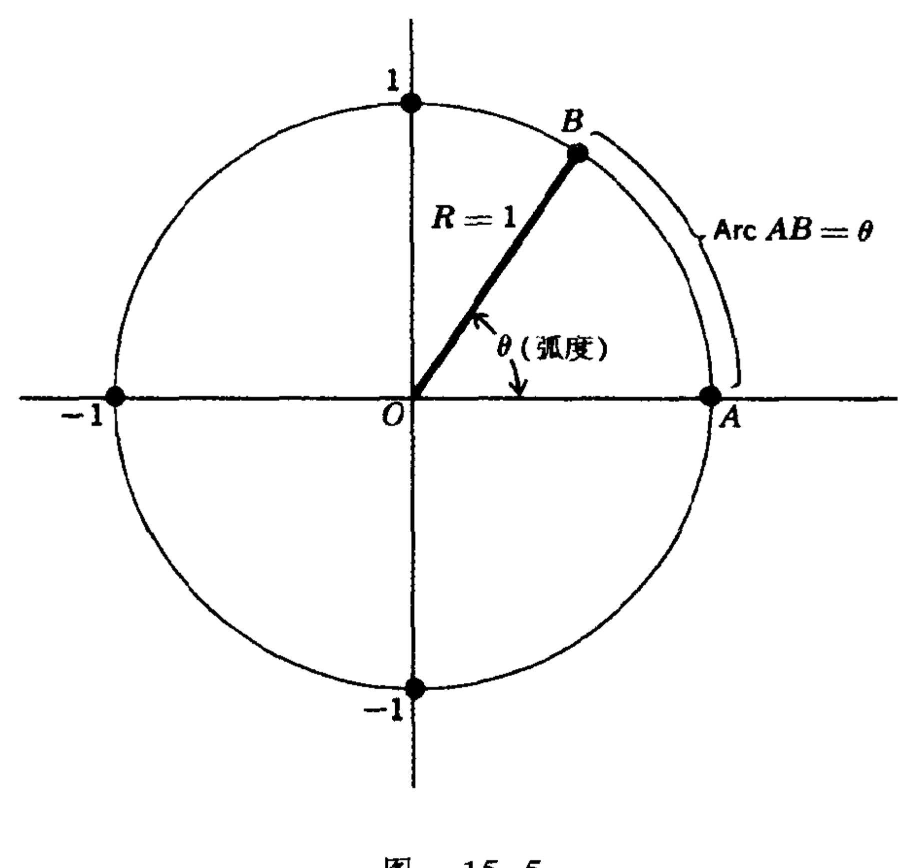



图 15.5

## 解法的一个例子

我们求微分方程

$$y''(t)+2y'(t)+17y=34$$

的解，其初始条件为 $y(0)=3, y'(0)=11$。

因为 $a_1=2, a_2=17, b=34$，马上可以求得特别积分为

$$y_p=\frac{b}{a_2}=\frac{34}{17}=2 \qquad [\text{由}(15.3)]$$

进而，因为 $a_1^2=4<4a_2=68$，所以，特征根为一对共轭复数 $(h\pm vi)$，其中

$$h = -\frac{1}{2}a_1 = -1 \quad 和 \quad v = \frac{1}{2}\sqrt{4a_2 - a_1^2} = \frac{1}{2}\sqrt{64} = 4$$

因此,由(15.24′),余函数为

$$y_c = e^{-t}(A_5\cos 4t + A_6\sin 4t)$$

合并 $y_c$ 和 $y_p$,可将通解表示成

$$y(t) = e^{-t}(A_5\cos 4t + A_6\sin 4t) + 2$$

我们运用两个初始条件来确定常数 $A_5$ 和 $A_6$。

首先,令通解中的 $t=0$,我们求得

526

$$\begin{aligned} y(0) &= e^0(A_5\cos 0 + A_6\sin 0) + 2 \\ &= (A_5 + 0) + 2 = A_5 + 2 \quad [\cos 0 = 1;\ \sin 0 = 0] \end{aligned}$$

运用初始条件 $y(0)=3$,我们可以确定 $A_5=1$。其次,将通解对 $t$ 微分[运用积的微分法则和导数公式(15.17)、(15.18),并记住链式法则(练习 15.2-5)],以求得 $y'(t)$,进而求 $y'(0)$:

$$y'(t) = -e^{-t}(A_5\cos 4t + A_6\sin 4t) + e^{-t}$$
$$[A_5(-4\sin 4t) + 4A_6\cos 4t]$$

从而

$$\begin{aligned} y'(0) &= -(A_5\cos 0 + A_6\sin 0) + (-4A_5\sin 0 + 4A_6\cos 0) \\ &= -(A_5 + 0) + (0 + 4A_6) = 4A_6 - A_5 \end{aligned}$$

运用第二个初始条件 $y'(0)=11$,并考虑到 $A_5=1$,则可知道 $A_6=3$。① 因此,定解为

$$y(t) = e^{-t}(\cos 4t + 3\sin 4t) + 2 \qquad (15.25)$$

同以前一样,$y_p(=2)$可以解释为 $y$ 的瞬时均衡水平,而 $y_c$ 则

① 注意,这里的 $A_6$ 实际上是一个实数,尽管在其定义中包含虚数 $i$。

表示与均衡水平的偏差。由于 $y_c$ 中存在三角函数，所以，可以预期时间路径(15.25)能够表现出波动形式。但它会包含何种具体波动形式呢？

**时间路径**

我们已经熟悉了图 15.4 所示的简单的正弦和余弦函数的路径。现在，我们必须研究正弦、余弦函数的某些变形和组合的路径，从而使我们能够一般地解释余函数(15.24′)

$$y_c = e^{ht}(A_5\cos vt + A_6\sin vt)$$

特别地，能够解释(15.25)的 $y_c$ 部分。

我们首先考察($A_5\cos vt$ )项。($\cos vt$)本身是 $vt$ 的三角函数，其周期为 $2\pi$，波幅为 1。周期为 $2\pi$ 意味着($vt$)每增加 $2\pi$，函数的图形就要重复一次。但当仅把 $t$ 视为自变量时，每当 $t$ 增加 $2\pi/v$ 时，图形便重复一次，所以相对于 $t$(在动态经济分析中是恰当的)，我们将($\cos vt$)的周期视为 $2\pi/v$。(但波幅仍为 1。)现在，当($\cos vt$)前附加乘积常数 $A_5$ 时，会导致波幅由 $\pm 1$ 变为 $\pm A_5$。因

527 此，尽管周期不受此常数的影响，但波幅现在却变为 $A_5$。同理，($A_6\sin vt$)是 $t$ 的正弦函数，周期为 $2\pi/v$，波幅为 $A_6$。

由于有相同的周期，和($A_5\cos vt + A_6\sin vt$)当 $t$ 增加 $2\pi/v$ 时，也会表现出重复的循环。为更严格地证明这一点，我们需注意，对于给定的 $A_5$ 和 $A_6$ 值，我们总可以求得两个常数 $A$ 和 $\varepsilon$，使得

$$A_5 = A\cos\varepsilon \quad 和 \quad A_6 = -A\sin\varepsilon$$

因此，我们可将所提到的和表示成

$$\begin{aligned} A_5\cos vt + A_6\sin vt &= A\cos\varepsilon\cos vt - A\sin\varepsilon\sin vt \\ &= A(\cos vt\cos\varepsilon - \sin vt\sin\varepsilon) \end{aligned}$$

$$= A\cos(vt+\varepsilon) \qquad [\text{由}(15.16)]$$

这是一个修正的 $t$ 的余弦函数,波幅为 $A$,周期为 $2\pi/v$,因为 $t$ 每增加 $2\pi/v$,$(vt+\varepsilon)$便增加 $2\pi$,在余弦曲线上便完成了一个波动周期。

如果 $y_c$ 仅由$(A_5\cos vt + A_6\sin vt)$所构成,则其含义应为:$y$ 的时间路径永无终结,围绕着 $y$ 的均衡值(以 $y_p$ 表示),以不变波幅波动。但实际上,还存在着一个乘积项 $e^{ht}$需要考虑。$e^{ht}$是非常重要的一项,因为正如我们将要看到的那样,它是决定时间路径是否收敛的关键。

若 $h>0$,则 $e^{ht}$的值将随 $t$ 的增加持续增加。这将对$(A_5\cos vt + A_6\sin vt)$的波幅产生放大效应,并依次在各个周期中导致与均衡的越来越大的偏离。如图 15.6$a$ 所示,在此情况下,时间路径具有**发散波动**的特征。另一方面,若 $h=0$,则 $e^{ht}=1$,余函数仅为$(A_5\cos vt + A_6\sin vt)$,我们已知道它有固定的波幅,在这种情况下,它具有如图 15.6$b$ 所示的时间路径,在每个周期中表现出与均衡均匀的偏离模式。所以,这是一个**均匀波动**的时间路径。最后,若 $h<0$,$e^{ht}$将会随 $t$ 的增加而持续递减,在依次各循环中,每个周期的波幅均比前一个周期的波幅要小,就像波纹逐渐消失一样。这种情形描绘在图 15.6$c$ 中,这种时间路径具有**衰减波动**的特征。(15.25)中的解,有 $h=-1$,是这种情况的一个例子。很清楚,只有这种衰减波动的情况才会产生**收敛**的时间路径;在其它两种情况下,时间路径是**非收敛**的或**发散**的。[1]

---

[1] 我们将交替使用**非收敛**和**发散**这两个词,尽管发散一词用于发散波动较之于非收敛的均匀波动要更严格一些。

528

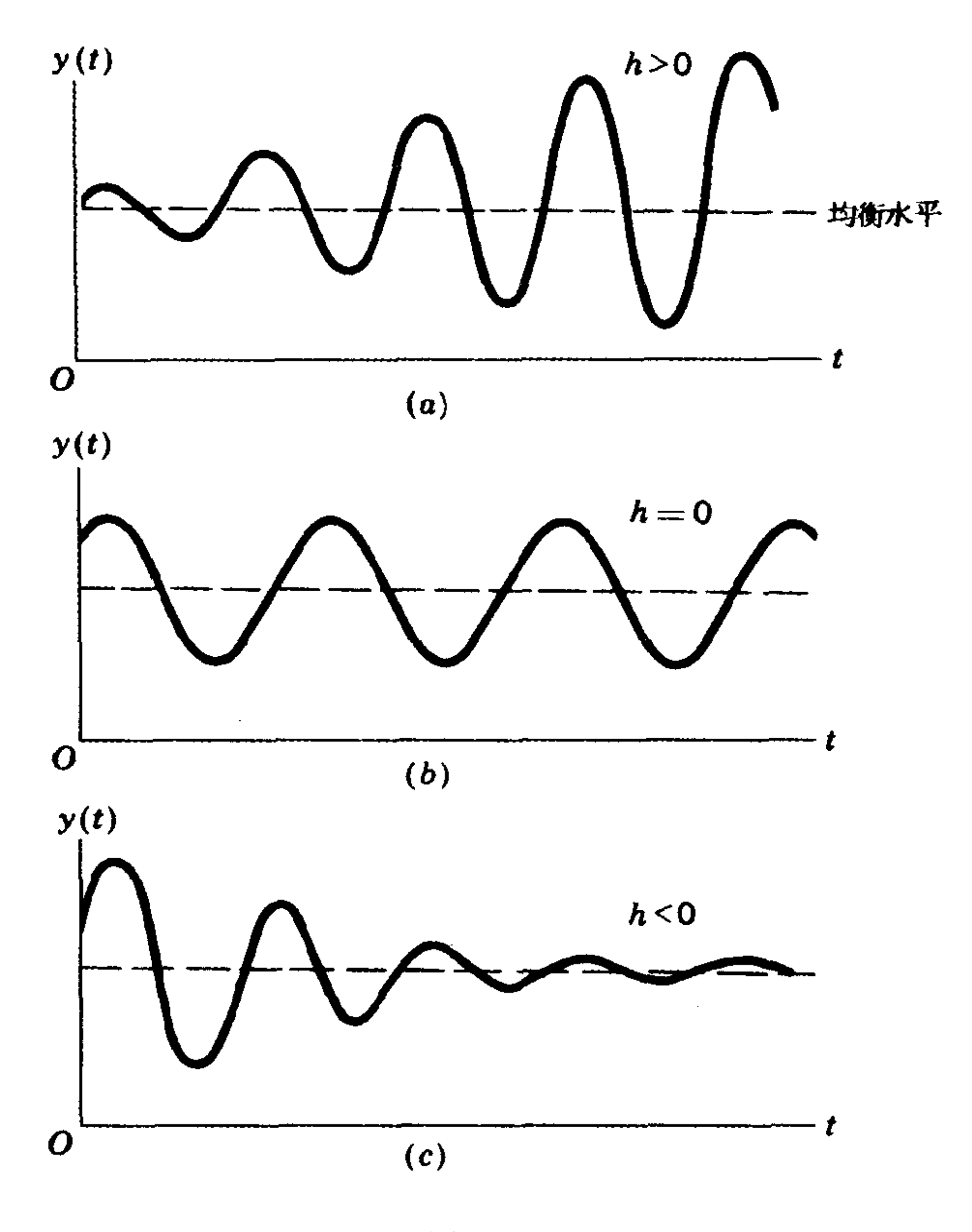



图　15.6

在图 15.6 中的三个图形中，假定瞬时均衡是静态的。如果瞬时均衡是移动均衡，那么，所描述的三种时间路径仍将围绕它波动，但因移动均衡一般描绘成一条曲线而非水平的直线，所以这种波动具有经济周期长期波动的性质。

## 均衡的动态稳定性

变量时间路径收敛性的概念与该变量瞬时均衡的动态稳定性

的概念有着密不可分的联系。具体地说，当且仅当时间路径是收敛的时，均衡才是动态稳定的。因此，$y(t)$路径收敛性的条件，即$h<0$（图 15.6$c$），也是 $y$ 的均衡的动态稳定性的条件。

读者可以回顾一下，对于特征根为实根的第一种和第二种情况，均衡动态稳定性的条件是每个特征根为负。在现在这种情况（第三种情况），由于具有复根，所以条件似乎更特殊一些，它只规定复根$(h \pm vi)$的实数部分$(h)$为负。但是，统一三种情况，并将看来不同的条件合并到一个一般的可应用的条件是可能的。只需 529 将任意实根 $r$ 看成一个虚数部分为零$(v=0)$的复根，便可以做到这一点。这样，条件"每个特征根的实数部分为负"，便可以应用于所有三种情况，并成为我们所需要的唯一条件。

## 练习 15.3

---

求下列每一方程的 $y_c$ 和 $y_p$、通解以及定解：

1 $y''(t)-4y'(t)+8y=0; y(0)=3, y'(0)=7$

2 $y''(t)+4y'(t)+8y=2; y(0)=2\frac{1}{4}, y'(0)=4$

3 $y''(t)+3y'(t)+4y=12; y(0)=2, y'(0)=2$

4 $y''(t)-2y'(t)+10y=5; y(0)=6, y'(0)=8\frac{1}{2}$

5 $y''(t)+9y=3; y(0)=1, y'(0)=3$

6 $2y''(t)-12y'(t)+20y=40; y(0)=4, y'(0)=5$

7 上述 6 个微分方程哪个能产生：($a$) 衰减波动；($b$) 均匀波动；($c$) 发散波动。

---

## 15.4 具有价格预期的市场模型

在前面形成的动态市场模型中，$Qd$ 和 $Qs$ 均只取为现期价格 $P$ 的函数。但是有时买者和卖者不仅将其市场行为建立在现期价格的基础上，而且建立在当时价格趋势的基础上。因为价格趋势可能使买者和卖者对未来价格做出某些预期，而这些价格预期又会影响其供求决策。

### 价格趋势与价格预期

在连续时间情况下，价格趋势信息基本上可由两个导数 $dP/dt$（价格是否上升）和 $d^2P/dt^2$（价格是否以递增速率上升）得到。为将价格趋势纳入考察范围，我们现在将这两个导数作为需求和供给函数的额外变量：

$$Q_d = D[P(t),\ P'(t),\ P''(t)]$$

$$Q_s = S[P(t),\ P'(t),\ P''(t)]$$

530 若我们仅限于讨论供求函数的线性形式，并将自变量简写为 $P$，$P'$ 和 $P''$，则我们可以写出

$$Q_d = \alpha - \beta P + mP' + nP'' \quad (\alpha, \beta > 0)$$

$$Q_s = -\gamma + \delta P + uP' + wP'' \quad (\gamma, \delta > 0) \qquad (15.26)$$

其中参数 $\alpha$、$\beta$、$\gamma$、$\delta$ 是从原来的市场模型带来的，而 $m$、$n$、$u$ 和 $w$ 是新的参数。

这四个尚未限定其符号的新参数，体现了买者和卖者的价格预期。例如，若 $m>0$，价格上升将导致 $Qd$ 增加。这也许表明，

买者预期价格会持续上升，因此，宁愿在现在价格相对较低时增加其购买。而若 $m<0$，表示买者的价格预期与前面正相反，所以他宁愿削减目前的购买而等待价格降低时再购进。函数中包含参数 $n$ 使得买者的行为也取决于 $dP/dt$ 的变化率。因此，新的参数 $m$ 和 $n$ 将价格投机这一因素纳入模型。从卖者的角度看，参数 $u$ 和 $w$ 也具有类似的含义。

## 一个简化的模型

为简化起见，我们假设需求函数仅含有价格预期因素。具体而言，令(15.26)中的 $m$ 和 $n$ 为非零，但令 $u=w=0$。进而假定在每一时点，市场均是出清的。然后，我们便可令需求和供给函数相等以得到(在正规化后)微分方程

$$P''+\frac{m}{n}P'-\frac{\beta+\delta}{n}P=-\frac{\alpha+\gamma}{n} \tag{15.27}$$

此方程为(15.2)经如下代换后的形式：

$$y=P \quad a_1=\frac{m}{n} \quad a_2=-\frac{\beta+\delta}{n} \quad b=-\frac{\alpha+\gamma}{n}$$

因为 $P$ 的变化模式涉及二阶导数 $P''$ 及一阶导数 $P'$，所以现在的模型显然有别于 14.2 节提出的动态市场模型。

但要注意，现在的模型与以前模型还有另一种不同。在 14.2 节，存在一种动态调节机制，$dP/dt=j(Qd-Qs)$。因为该方程意味着当且仅当 $Qd=Qs$ 时，$dP/dt=0$，所以在该模型中，瞬时意义上的均衡和市场出清意义上的均衡是一致的。与此相反，现在的模型假设在每一时刻市场均是出清的，因此，市场所达到的每一个价格均是市场出清意义上的均衡价格。换言之，两种意义的均衡现在是分离的。但也要注意，调节机制 $dP/dt=j(Qd-Qs)$ 含有一个导

数，正是它使前面的市场模型成为动态模型。在现在的模型中不具
531 有调节机制，模型的动态性质源于预期项 $mP'$ 和 $nP''$。

### 价格的时间路径

此模型的瞬时均衡价格——特别积分 $P_p$(前面记为 $y_p$)——很容易用(15.3)求得。它是

$$P_p = \frac{b}{a_2} = \frac{\alpha + \gamma}{\beta + \delta}$$

因为特别积分是一个正常数，所以它表示一个稳定均衡。

至于余函数 $P_c$(前面记为 $y_p$)，存在三种可能的情况。

**第一种情况(不同的实根)**

$$\left(\frac{m}{n}\right)^2 > -4\left(\frac{\beta + \delta}{n}\right)$$

由(15.7)，这种情况的余函数为

$$P_c = A_1 e^{r_1 t} + A_2 e^{r_2 t}$$

其中

$$r_1, r_2 = \frac{1}{2}\left[-\frac{m}{n} \pm \sqrt{\left(\frac{m}{n}\right)^2 + 4\left(\frac{\beta + \delta}{n}\right)}\right] \quad (15.28)$$

相应地，通解为

$$P(t) = P_c + P_p = A_1 e^{r_1 t} + A_2 e^{r_2 t} + \frac{\alpha + \gamma}{\beta + \delta} \quad (15.29)$$

**第二种情况(二重实根)**

$$\left(\frac{m}{n}\right)^2 = -4\left(\frac{\beta + \delta}{n}\right)$$

在此情况下，特征根取一个值

$$r = -\frac{m}{2n}$$

因此,由(15.9),通解可以写成

$$P(t) = A_3e^{-mt/2n} + A_4te^{-mt/2n} + \frac{\alpha+\gamma}{\beta+\delta} \quad (15.29')$$

**第三种情况(复根)**

$$\left(\frac{m}{n}\right)^2 < -4\left(\frac{\beta+\delta}{n}\right)$$

在第三种情况下,特征根是一对共轭复数。

$$r_1,\ r_2 = h \pm vi$$

其中

532

$$h = -\frac{m}{2n} \quad 和 \quad v = \frac{1}{2}\sqrt{-4\left(\frac{\beta+\delta}{n}\right)-\left(\frac{m}{n}\right)^2}$$

因此,由(15.24′),我们有通解

$$P(t) = e^{-mt/2n}(A_5\cos vt + A_6\sin vt) + \frac{\alpha+\gamma}{\beta+\delta} \quad (15.29'')$$

可由这些结果推出一对通解。首先,若 $n>0$,则 $-4(\beta+\delta)/n$ 必定为负,因而小于$(m/n)^2$。所以可立即排除第二、三种情况。进而,由于 $n$ 为正(同 $\beta$ 和 $\delta$ 一样),所以(15.28)中平方根号下的表达式必然大于$(m/n)^2$,因此,其平方根必然大于$|m/n|$。那么,(15.28)中的±号会产生一个正根($r_1$)和一个负根($r_2$)。结果,瞬时均衡是动态不稳定的,除非(15.29)中确定了的常数 $A_1$ 值恰好为零。

其次,若 $n<0$,则三种情况均是可行的。在第一种情况下,若 $m$ 为负,则我们可以确定两个根均为负(为什么?)。有趣的是,如果 $m$ 为负,第二种情况下的重根也将为负。而且,因为第三种情况下

的复根的实数部分 $h$，与第二种情况下的重根$\gamma$ 取相同的值，所以，$m$ 为负也可以保证 $h$ 为负。总之，对于所有这三种情况，当参数 $m$ 和 $n$ 均为负时，均衡的动态稳定性是可以保证的。

**例1** 令需求与供给函数为

$$Q_d = 42 - 4P - 4P' + P''$$

$$Q_s = -6 + 8P$$

初始条件为 $P(0)=6, P'(0)=4$。假设在每一时点市场均是出清的，求时间路径 $P(t)$。

在本例中，参数值为

$$\alpha = 42 \quad \beta = 4 \quad \gamma = 6 \quad \delta = 8 \quad m = -4 \quad n = 1$$

因为 $n$ 为正，所以我们前面的讨论表明，只有第一种情况可以产生，且两个(实)根 $r_1$ 和 $r_2$ 均将为正。将参数值代入(15.28)的确可以证实这一点，因为

$$r_1, \; r_2 = \frac{1}{2}(4 \pm \sqrt{16+48}) = \frac{1}{2}(4 \pm 8) = 6, \; -2$$

则由(15.29)，通解为

$$P(t) = A_1 e^{6t} + A_2 e^{-2t} + 4$$

进而，考虑到初始条件，求得 $A_1 = A_2 = 1$，从而定解为

$$P(t) = e^{6t} + e^{-2t} + 4$$

由于正根 $r_1 = 6$，所以瞬时均衡($P_p = 4$)是动态不稳定的。

533 上面的解是运用公式(15.28)、(15.29)求出的。此外，还可以首先令给定需求函数和供给函数相等，以得到微分方程。

$$P'' - 4P' - 12P = -48$$

然后像(15.2)那种情况一样解此方程，也可以得到解。

**例2** 已知需求函数和供给函数

$$Q_d = 40 - 2P - 2P' - P''$$

$$Q_s = -5 + 3P$$

初始条件 $P(0)=12, P'(0)=1$,在市场总为出清的假设条件下求 $P(t)$。

这里,参数 $m$ 和 $n$ 均为负。因此,按照前面的一般论述,瞬时均衡应为动态稳定的。为求出具体解,我们可以首先令 $Qd$ 与 $Qs$ 相等以得到如下微分方程(在以 $-1$ 通乘后)

$$P'' + 2P' + 5P = 45$$

由特别积分给出瞬时均衡

$$P_p = \frac{45}{5} = 9$$

由微分方程的特征方程

$$r^2 + 2r + 5 = 0$$

可以求得复根

$$r_1, r_2 = \frac{1}{2}(-2 \pm \sqrt{4-20}) = \frac{1}{2}(-2 \pm 4i) = -1 \pm 2i$$

这意味着 $h=-1, v=2$,从而通解为

$$P(t) = e^{-t}(A_5\cos 2t + A_6\sin 2t) + 9$$

为确定任意常数 $A_5$ 和 $A_6$,令通解中的 $t=0$,得到

$$P(0) = e^0(A_5\cos 0 + A_6\sin 0) + 9$$

$$= A_5 + 9 \quad [\cos 0 = 1;\ \sin 0 = 0]$$

进而,微分通解,然后令 $t=0$,求得

534

$$P'(t) = -e^{-t}(A_5\cos 2t + A_6\sin 2t) + e^{-t}$$

$$(-2A_5\sin 2t + 2A_6\cos 2t)$$

[积的求导法则和链式法则]

和 $P'(0) = -e^0(A_5\cos 0 + A_6\sin 0)$

$$+ e^0(-2A_5\sin 0 + 2A_6\cos 0)$$

$$= -(A_5 + 0) + (0 + 2A_6) = -A_5 + 2A_6$$

因此，利用初始条件 $P(0)=12, P'(0)=1$，有 $A_5=3, A_6=2$。因而，定解为

$$P(t) = e^{-t}(3\cos 2t + 2\sin 2t) + 9$$

这个时间路径显然是一个周期波动的时间路径，周期为$2\pi/v=\pi$。即$t$每增加$\pi=3.14159\cdots\cdots$，便存在一个完整的周期。考虑到乘积项$e^{-t}$，可知波动是衰减的。始于初始价格$P(0)=12$的时间路径，以循环方式收敛于瞬时均衡价格$P_p=9$。

## 练习 15.4

---

1 令(15.26)中的参数 $m$、$n$、$u$ 和 $w$ 均不为零。

($a$) 假设在每一时点市场均出清，写出该模型的新的微分方程。

($b$) 求瞬时均衡价格。

($c$) 在什么条件下，可以排除周期波动？

2 令需求函数和供给函数与(15.26)中的需求函数和供给函数相同，且同文中讨论的一样，有 $u=w=0$。

($a$) 若市场并非总是出清的，而是按

$$\frac{dP}{dt} = j(Q_d - Q_s) \quad (j > 0)$$

进行调整，写出新的恰当的微分方程。

($b$) 求瞬时均衡价格和市场出清均衡价格。

($c$) 写出波动价格路径的条件。在 $n>0$ 情况下能发生波动吗？

3 令需求函数和供给函数为

$$Q_d = 9 - P + P' + 3P'' \quad Q_s = -1 + 4P - P' + 5P''$$

且有 $P(0)=4$ 及 $P'(0)=4$。

($a$) 假设市场在每一时点都出清,求价格路径。

($b$) 时间路径收敛吗? 具有波动吗?

---

# 15.5 通货膨胀与失业的相互作用

本节介绍应用二阶微分方程的一个宏观模型,该模型涉及通货膨胀与失业问题。535

## 菲利普斯关系

在通货膨胀与失业问题的现代分析中,最广泛使用的一个概念是菲利普斯关系。[①] 菲利普斯原来的公式是描述货币工资率与失业率之间的负的经验关系:

$$w = f(U) \qquad [f'(U) < 0] \qquad (15.30)$$

其中小写字母 $w$ 表示货币工资 $W$ 的增长率(即 $w=(\dot{W}/W)$, $U$ 表示失业率。所以,它仅与劳动市场有关。但在后来的应用中,已将菲利普斯关系调整为一种将通货膨胀率(而非 $w$)与失业率联系起来的函数。这种调整或许以这种观点为基础:价格上升是普遍存在的,从而反映增长的货币工资成本的 $w$ 为正,这必然带有通货膨胀的含义。而这又使得通货膨胀率像 $w$ 一样,是 $U$ 的函数。但正 $w$ 的膨胀性压力可能被(假定为外生并以 $T$ 表示的)劳动生产率的增长所抵消。具体而言,膨胀性效应仅当货币工资增

---

① A.W.菲利普斯:"1861—1957 年英国的失业与货币工资变化率间的关系",载《经济学》,1958 年 11 月,283—299 页。

长快于生产率增长时方能具体表现出来。以小写字母 $p$ 表示通货膨胀率(即价格水平 $P$ 的增长率,$p=\dot{P}/P$),则可写出

$$p = w - T \tag{15.31}$$

合并(15.30)和(15.31),并采用 $f(U)$ 函数的线性形式,则我们可以得到调整的菲利普斯关系:

$$p = \alpha - T - \beta U \quad (\alpha, \beta > 0) \tag{15.32}$$

**预期增加的菲利普斯关系**

最近,经济学家喜欢采用预期增加的菲利普斯关系:

$$w = f(U) + h\pi \quad (0 < h \leqslant 1) \tag{15.30'}$$

其中$\pi$ 表示预期的通货膨胀率。正如弗里德曼教授所描述的那样,[①] (15.30′)所包含的思想是:如果膨胀性趋向在相当长时期内存在,人们便会形成某种通货膨胀预期,并力图将这种预期纳入其货币工资需求。因此,$w$ 应为$\pi$ 的增函数。将这种思想纳入到(15.32)中,产生如下方程

536

$$p = \alpha - T - \beta U + h\pi \quad (0 < h \leqslant 1) \tag{15.33}$$

我们已引入一个新的变量来表示预期的通货膨胀,所以有必要假设通货膨胀预期是如何具体形成的。[②] 我们这里采用适应性预期假设:

$$\frac{d\pi}{dt} = j(p - \pi) \quad (0 < j \leqslant 1) \tag{15.34}$$

---

① 米尔顿·弗里德曼:"货币政策的作用",载《美国经济评论》,1968 年 3 月号,1—17页。

② 这与上一节相反。在上一节讨论价格预期时,未引入新的变量来表示预期价格。因而,关于预期形成的假设暗含于(15.26)中的参数 $m$、$n$、$u$ 和 $w$ 中。

注意，此方程并不解释$\pi$的绝对大小，而是描述其随时间变化的方式。如果实际通货膨胀率 $p$ 超过预期通货膨胀率$\pi$，那么，现在已证明是过低的$\pi$将会向上调整($d\pi/dt>0$)。反之，若 $p$ 低于$\pi$，则$\pi$就会向下调整。在形式上，(15.34)非常类似于市场模型的调节机制 $dP/dt=j(Qd-Qs)$。但在这里，驱动调整的力量是实际通货膨胀与预期通货膨胀的偏差，而非 $Qd$ 与 $Qs$ 的偏差。

## 从通货膨胀到失业的反馈

可以认为，(15.33)和(15.34)构成了一个完整的模型。但是，由于在两个方程中存在三个变量，所以，其中的一个变量必须视为外生的。比如，若我们将$\pi$和 $p$ 视为内生的，则必须将 $U$ 视为外生的。还有一个更满意的选择是引入第三个方程来解释变量 $U$，这样，模型会包含更丰富的行为特征。更重要的是，这将为我们提供一个考虑通货膨胀对失业的反馈效果的机会。方程(15.33)告诉我们 $U$ 如何影响 $p$(主要从该经济的供给方面考察)。但 $p$ 无疑又会影响 $U$。例如，通货膨胀率可能会影响公众的消费储蓄决策，因而影响到对国内产品的总需求，而这又会影响到失业率。甚至在政府需求管理政策的指导下，通货膨胀率也会使得政策效果产生不同。在不同的通货膨胀率条件下，一个给定的货币支出水平(财政政策)可能会转化为不同的实际支出水平，同样，一个给定的名义货币扩张(货币政策)可能产生不同的实际货币扩张率。而这些，又会对产出和失业产生不同的影响。 537

为简便计，我们仅考察通过货币政策传递的反馈。以 $M$ 表示名义货币余额，名义货币余额的增长率以 $m\equiv\dot{M}/M$ 来表示，

我们假设[1]

$$\frac{dU}{dt} = -k(m-p) \quad (k > 0) \tag{15.35}$$

回忆一下(10.25),并在这里应用,我们知道($m-p$)式表示实际货币增长率:

$$m - p = \frac{\dot{M}}{M} - \frac{\dot{p}}{p} = r_M - r_P = r_{(M/P)}$$

因此,(15.35)规定了 $dU/dt$ 与实际货币余额增长率负相关。由于变量 $p$ 现在成为 $dU/dt$ 的一个决定因素,所以模型现在包含了从通货膨胀到失业的反馈。

## $\pi$ 的时间路径

(15.33),(15.34)及(15.35)构成了一个包含三个变量 $\pi$、$p$ 和 $U$ 的封闭模型。但是,消去其中的两个变量,我们可以将模型化简为一个单变量的微分方程。假设我们令单变量为 $\pi$,则可首先将(15.33)代入(15.34)得到

$$\frac{d\pi}{dt} = j(\alpha - T - \beta U) - j(1-h)\pi \tag{15.36}$$

如果此方程含有表达式 $dU/dt$ 而非 $U$,我们可以直接代入(15.35)。但正如(15.36)所表明的那样,我们必须首先将(15.36)对 $t$ 微分以得到 $dU/dt$ 项:

$$\frac{d^2\pi}{dt^2} = -j\beta\frac{dU}{dt} - j(1-h)\frac{d\pi}{dt} \tag{15.37}$$

将(15.35)代入此式,则产生

---

① 在前面的讨论中,我们曾以 $M_s$ 表示货币供给,使之与货币需求 $M_d$ 相区别。在这里,我们可以直接使用不带下标的字母 $M$ 而不必担心造成混淆。

$$\frac{d^2\pi}{dt^2} = j\beta \mathrm{km} - j\beta \mathrm{kp} - j(1-h)\frac{d\pi}{dt} \qquad (15.37')$$

这里仍有一个变量 $p$ 有待于消去。为消去它，我们注意到，(15.34)意味着

$$p = \frac{1}{j}\frac{d\pi}{dt} + \pi \qquad (15.38)$$

将此结果代入(15.37′)，化简，便可得到所求的仅含变量$\pi$ 的微分方程 538

$$\frac{d^2\pi}{dt^2} + \underbrace{[\beta k + j(1-h)]}_{a_1}\frac{d\pi}{dt} + \underbrace{(j\beta \mathrm{k})}_{a_2}\pi = \underbrace{j\beta \mathrm{km}}_{b} \qquad (15.37'')$$

此方程的特别积分为

$$\pi_p = \frac{b}{a_2} = m$$

因此，预期通货膨胀率的瞬时均衡值只取决于名义货币的增长率。

对于余函数，同前面一样，两个根为

$$r_1,\ r_2 = \frac{1}{2}\left(-a_1 \pm \sqrt{a_1^2 - 4a_2}\right) \qquad (15.39)$$

其中，$a_1$ 和 $a_2$ 均为正[在(15.37″)就应注明]，由于事先不可能确定 $a_1^2$ 是大于，等于，还是小于 $4a_2$，所以特征根三种可能的情况——不同的实根、重实根、复根——均有可能产生。然而，无论出现哪种情况，在现在的模型中，瞬时均衡总是动态稳定的。这是因为：首先，假设出现第一种情况，有 $a_1^2 > 4a_2$，则(15.39)中的平方根会产生一个实数。因为 $a_2$ 为正，$\sqrt{a_1^2 - 4a_2}$必然小于 $\sqrt{a_1^2} = a_1$。由此知 $r_1$ 为负，$r_2$ 也为负，这意味着动态稳定均衡。若 $a_1^2 = 4a_2$(第二种情况)会如何呢？在此情况下，平方根为零，从而 $r_1 =$

$r_2 = -a_1/2 < 0$。负重根同样意味着动态稳定性。最后,对于第三种情况,复根的实数部分为 $h = -a_1/2$。因为此值与第二种情况下的重根相同,所以,关于动态稳定性会得出同样的结论。

尽管我们只研究了$\pi$ 的时间路径,但模型无疑也会得出其它变量的信息。比如,为求出变量 $U$ 的时间路径,我们既可以将模型化为仅含 $U$ 而非$\pi$ 的微分方程(参见练习 15.5-2),也可以由早已求出的$\pi$ 的路径推导出 $U$ 路径(参见例 1)。

**例 1** 令模型的三个方程取如下具体形式

$$p = \frac{1}{6} - 3U + \pi \tag{15.40}$$

$$\frac{d\pi}{dt} = \frac{3}{4}(p - \pi) \tag{15.41}$$

$$\frac{dU}{dt} = -\frac{1}{2}(m - p) \tag{15.42}$$

则我们有参数值 $\beta = 3, h = 1, j = 3/4, k = 1/2$。因此,参照(15.
539 37″),求得

$$a_1 = \beta k + j(1-h) = \frac{3}{2} \quad a_2 = j\beta k = \frac{9}{8} \text{ 和 } b = j\beta k m = \frac{9}{8}m$$

特别积分为 $b/a_2 = m$。有 $a_1^2 < 4a_2$,所以特征根为复根:

$$r_1,\ r_2 = \frac{1}{2}\left(-\frac{3}{2} \pm \sqrt{\frac{9}{4} - \frac{9}{2}}\right)$$

$$= \frac{1}{2}\left(-\frac{3}{2} \pm \frac{3}{2}i\right) = -\frac{3}{4} \pm \frac{3}{4}i$$

即 $h = -3/4, v = 3/4$。结果,预期通货膨胀率的通解为

$$\pi(t) = e^{-3t/4}\left(A_5\cos\frac{3}{4}t + A_6\sin\frac{3}{4}t\right) + m \tag{15.43}$$

它描述了一个围绕均衡值 $m$ 衰减波动的时间路径。

由此,我们还可以推出变量 $p$ 和 $U$ 的时间路径。根据(15.41),$p$ 可以表示成$\pi$ 和 $d\pi/dt$ 的方程

$$p = \frac{4}{3}\frac{d\pi}{dt} + \pi$$

通解(15.43)的$\pi$ 路径意味着导数

$$\frac{d\pi}{dt} = -\frac{3}{4}e^{-3t/4}\left(A_5\cos\frac{3}{4}t + A_6\sin\frac{3}{4}t\right) + e^{-3t/4}\left(-\frac{3}{4}A_5\sin\frac{3}{4}t + \frac{3}{4}A_6\cos\frac{3}{4}t\right)$$

[积的微分法则和链式法则]

用(15.43)的解及其导数,我们可以推出

$$p(t) = e^{-3t/4}(A_6\cos\frac{3}{4}t - A_5\sin\frac{3}{4}t) + m \qquad (15.44)$$

同预期的通货膨胀率$\pi$ 一样,实际通货膨胀率 $p$ 也具有收敛于均衡值 $m$ 的波动的时间路径。

至于变量 $U$,由(15.40)可知,它可以按$\pi$ 和 $p$ 表示如下:

$$U = \frac{1}{3}(\pi - p) + \frac{1}{18}$$

因此,利用解(15.43)和(15.44),我们可将失业的时间路径写成

$$U(t) = \frac{1}{3}e^{-3t/4}\left[(A_5 - A_6)\cos\frac{3}{4}t + (A_5 + A_6)\sin\frac{3}{4}t\right] + \frac{1}{18} \qquad (15.45)$$

此路径同样是一个衰减波动路径。$U$ 的动态稳定瞬时均衡值为 1/18。

由于$\pi$ 和 $p$ 的瞬时均衡值都等于货币政策参数 $m$,$m$ 的值(名义货币增长率)提供了一个轴,$\pi$和$p$的时间路径均围绕此轴 540

波动。如果 $m$ 发生变化,那么,$\pi$ 和 $p$ 的新的均衡值会立即取代原来的值,且无论$\pi$ 变量和 $p$ 变量在货币政策变化时恰好取何值,都将成为新的$\pi$ 和 $p$ 路径产生的初始值。

相反,$U$ 的瞬时均衡值并不依赖于 $m$。根据(15.45),无论名义货币增长率为多少,也无论均衡通货膨胀率为多少,$U$ 都收敛于常数 1/18。这个不变的 $U$ 的均衡值称作自然失业率。自然失业率与任何均衡通货膨胀率相一致这一事实,可以在 $Up$ 空间中以一条平行于 $p$ 轴的纵直线来表示。此垂线将 $U$ 和 $p$ 的均衡值联系起来,称作长期菲利普斯曲线。但是这条曲线的垂直形状,是由本例特定的参数值所决定的。当这些参数值改变时,如像 15.5-4 那样,长期菲利普斯曲线也许不再是垂直的。

## 练习 15.5

---

1 在通货膨胀和失业模型中,保留(15.33)和(15.34),但除去(15.35),并令 $U$ 为外生的。

($a$) 现在会产生何种微分方程?

($b$) 会得到几个特征根?现在余函数还可能具有周期性波动的特征吗?

2 在课文的讨论中,我们将通货膨胀失业模型化为以$\pi$ 为变量的微分方程。证明:模型还可以化为一个变量 $U$ 的二阶微分方程,此方程与(15.37″)具有同样的系数 $a_1$ 和 $a_2$,但有不同的常数项 $b=kj[\alpha-T-(1-h)m]$。

3 令适应性预期假设(15.34)为所谓的"完美预期"假设$\pi=p$ 所代替,保留(15.33)和(15.35)。

($a$) 导出以 $p$ 为变量的微分方程。

($b$) 导出以 $U$ 为变量的微分方程。

($c$) 这些方程与我们在适应性预期假设条件下所得到的方程有何根本不同?

($d$) 为使新的微分方程有意义,对参数的限制必须作何变化?

4　在例 1 中,保留(15.41)和(15.42),但以下式代替(15.40)

$$p = \frac{1}{6} - 3U + \frac{1}{3}\pi$$

($a$) 求 $p(t)$,$\pi(t)$和 $U(t)$。

($b$) 这些时间路径仍为波动的吗? 仍收敛吗?

($c$) $p$ 和 $U$ 的瞬时均衡值 $\bar{p}$ 和 $\bar{U}$ 为多少?

541

($d$) $\bar{U}$ 和 $\bar{p}$ 仍不存在函数关系吗? 若我们现在用长期菲利普斯曲线将两个均衡值联系起来,我们仍能得到一条垂线吗? 那么,例 1 中的哪个假设对导出垂直的长期菲利普斯曲线有决定作用?

---

## 15.6　具有可变项的微分方程

在上面考察的微分方程

$$y''(t) + a_1 y'(t) + a_2 y = b$$

中,右边的那项 $b$ 是一个常数。若该项不是 $b$,而是一个可变项,即 $t$ 的某个函数,如 $bt^2$, $e^{bt}$,或 $b\sin t$,那么,会如何呢? 答案是,那时我们必须修正我们的特别积分 $y_p$。幸运的是,余函数不因可变项的存在而受影响,因为 $y_c$ 仅涉及右边始终为零的简化方程。

### 待定系数法

我们将介绍一种求 $y_p$ 的方法。这种方法称作待定系数法;只要变项及其逐次导数仅含有有限的不同的表达式(乘积常数除外),就可以应用它去解常系数可变项的微分方程。我们最好通过

具体例子来说明这一点。

**例 1** 求微分方程

$$y''(t) + 5y'(t) + 3y = 6t^2 - t - 1 \qquad (15.46)$$

的特别积分。根据定义，特别积分是满足给定方程的 $y$ 值，即无论 $t$ 为何值，均使方程的左边恒等于右边的 $y$ 值。因为方程的左边包含函数 $y(t)$ 和导数 $y'(t)$ 及 $y''(t)$——而右边含有 $t^2$ 及 $t$ 的倍数、常数——我们问：一般函数 $y(t)$ 取何种形式，与其一阶、二阶导数一道，会给出三种类型的表达式 $t^2$、$t$ 和一个常数呢？显然，答案是形式为 $B_1t^2 + B_2t + B_3$（其中 $B_i$ 也为待定系数）的函数，因为若我们将特别积分写成

$$y(t) = B_1t^2 + B_2t + B_3$$

可以导出

$$y'(t) = 2B_1t + B_2 \text{ 和 } y''(t) = 2B_1 \qquad (15.47)$$

542 这三个方程确实包含所说的表达式的类型。将其代入(15.46)并合并各项，得到

$$\text{左边} = (3B_1)t^2 + (10B_1 + 3B_2)t + (2B_1 + 5B_2 + 3B_3)$$

当此式与(15.46)右边的各项逐项相等时，我们可确定系数 $B_i$ 如下：

$$\left.\begin{aligned} 3B_1 &= 6 \\ 10B_1 + 3B2 &= -1 \\ 2B_1 + 5B_2 + 3B_3 &= -1 \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} B_1 &= 2 \\ B_2 &= -7 \\ B_3 &= 10 \end{aligned}\right.$$

因此，所求特别积分可以写成

$$y_p = 2t^2 - 7t + 10$$

仅当表达式的类型为有限个时，方能应用这种方法。（参见练

习 15.6 - 1。)一般而言,当此前提条件满足时,可以视特别积分为包含所有给定可变项及其导数的所有不同类型表达式的线性组合。特别要注意,如果原可变项及其各阶导数含有常数项,那么,特别积分中也应包括常数项。

**例 2** 作为进一步的解释说明,我们来求适于可变项($b\sin t$)的特别积分的一般形式,在此情况下,反复微分会产生逐阶导数($b\cos t$),($-b\sin t$),($-b\cos t$),($b\sin t$)等等,它们仅包含两种类型的表达式。因此,我们可以试验形式为($B_1\sin t + B_2\cos t$)的特别积分。

## 一个修正

在某些情况下,应用这种方法会产生一些复杂情况。当给定微分方程中 $y$ 项的系数为零,比如像下式那样,

$$y''(t) + 5y'(t) = 6t^2 - t - 1$$

前面使用的 $y_p$ 的试探形式,即 $B_1t^2 + B_2t + B_3$ 不能够再应用。导致其失效的原因在于 $y(t)$项不存在,因而如(15.47)所示,只有 $y'(t)$和 $y''(t)$可代入方程左边,所以方程左边没有 $B_1t^2$ 项与方程右边 $6t^2$ 项相等。排除这一困难的方法是运用试探解 $t(B_1t^2 + B_2t + B_3)$;若此试探解也失效,则使用 $t^2(B_1t^2 + B_2t + B_3)$,如此等等。

同样的方法还可以应用于下例所描述的另一种困难的环境。

**例 3** 求

543

$$y''(t) + 3y'(t) - 4y = 2e^{-4t} \tag{15.48}$$

的特别积分。这里,可变项的形式为 $e^{-4t}$,而其各阶导数,即 $-8e^{-4t}$,$32e^{-4t}$,$-128e^{-4t}$等,也取同样的形式。若我们尝试解

$$y(t) = Be^{-4t}$$

$$[\text{具有 } y'(t) = -4Be^{-4t} \quad \text{和} \quad y''(t) = 16Be^{-4t}]$$

并将其代入(15.48)，得到一个不太理想的结果

$$\text{左边} = (16-12-4)Be^{-4t} = 0 \tag{15.49}$$

它显然不等于方程右边的项 $2e^{-4t}$。

导致这一结果的原因是，可变项的指数系数恰好是(15.48)的特征方程的一个根：

$$r^2 + 3r - 4 = 0 \qquad (\text{根 } r_1, r_2 = 1, -4)$$

回顾一下，特征方程是通过微分过程得到的。[①] 而(15.49)中的式(16 - 12 - 4)也是通过同样的过程导出的。因此，毫不奇怪，(16 - 12 - 4)只不过是($r^2 + 3r - 4$)当 $r = -4$ 的一种特殊形式。因为 -4 恰好为特征根，所以二次表达式

$$r^2 + 3r - 4 = 16 - 12 - 4$$

必然恒为零。

为克服这一问题，我们尝试解

$$y(t) = Bte^{-4t}$$

它具有导数

$$y'(t) = (1-4t)Be^{-4t} \text{ 和 } y''(t) = (-8+16t)Be^{-4t}$$

将其代入(15.48)会得到：方程左边 = $-5Be^{-4t}$。当它与方程右边相等时，可确定系数为 $B = -2/5$。因而，所求的(15.48)的特别积分为

$$y_p = \frac{-2}{5}te^{-4t}$$

① 参见导出(15.4″)的讨论。

## 练习 15.6

---

1 证明:待定系数法不能应用于微分方程 $y''(t)+ay'(t)+by=t^{-1}$。

2 运用待定系数法求下列每一方程的特别积分: 544

(a) $y''(t)+2y'(t)+y=t$

(b) $y''(t)+4y'(t)+y=2t^2$

(c) $y''(t)+y'(t)+2y=e^t$

(d) $y''(t)+y'(t)+3y=\sin t$

---

# 15.7 高阶线性微分方程

上面介绍的解法很容易推广到 $n$ 阶线性微分方程的情况。具有常系数和常数项的微分方程一般可以写成

$$y^{(n)}(t)+a_1y^{(n-1)}(t)+\cdots+a_{n-1}y'(t)+a_ny=b \tag{15.50}$$

## 求 解

在具有常系数和常数项的情况下,高阶导数的存在对上面所讨论的求特别积分的方法并无实质影响。

若我们采用最简单的试探解 $y=k$,可以看到,从 $y'(t)$ 到 $y^{(n)}(t)$ 的所有导数均为零。因此,(15.50)简化为 $a_nk=b$,且我们可以写出

$$y_p=k=\frac{b}{a_n}\ (a_n\neq 0) \qquad [参见\ (15.3)]$$

但是，在 $a_n=0$ 的情况下，我们必须尝试形式为 $y=kt$ 的解。则因 $y'(t)=k$，而所有更高阶导数将为零，所以，(15.50)可以简化为 $a_{n-1}k=b$，因而产生特别积分

$$y_p = kt = \frac{b}{a_{n-1}}t \quad (a_n = 0;\ a_{n-1} \neq 0) \qquad [\text{参阅 } (15.3')]$$

若恰好出现 $a_n=a_{n-1}=0$，则上面这个解也将无效。因而必须试探形式为 $y=kt^2$ 的解。这个方法的进一步修正应是很显然的。

至于余函数，微分方程中包含高阶导数会产生提高特征方程次数的效果。余函数被定义为简化方程

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_{n-1}y'(t) + a_n y = 0 \tag{15.51}$$

545 的通解。以 $y=Ae^{rt}(\neq 0)$ 作为试探解，并利用 $y'(t)=rAe^{rt}$，$y''(t)=r^2Ae^{rt},\cdots,y^{(n)}(t)=r^nAe^{rt}$，可将(15.51)重写成

$$Ae^{rt}(r^n + a_1 r^{n-1} + \cdots + a_{n-1}r + a_n) = 0$$

满足如下($n$ 次多项式)特征方程

$$r^n + a_1 r^{n-1} + \cdots + a_{n-1}r + a_n = 0 \tag{15.51'}$$

的任意 $r$ 值，也满足上面(15.51)的改写形式，当然，此多项式方程有 $n$ 个根，其中每个根均应包含于(15.51)的通解中。因此，我们的余函数的一般形式为

$$y_c = A_1 e^{r_1 t} + A_2 e^{r_2 t} + \cdots + A_n e^{r_n t} \quad \left(= \sum_{i=1}^{n} A_i e^{r_i t}\right)$$

但同以前一样，在 $n$ 个根并不都是实根及不同的根的情况下，必须作某些修正。首先，假设存在重根，比如 $r_1=r_2=r_3$，则为避免“重叠”，我们必须将解的前三项写成 $A_1e^{r_1t}+A_2te^{r_1t}+A_3t^2e^{r_1t}$[参见(15.9)]。在我们有 $r_4=r_1$ 的情况下，第四项必须

调整为 $A_4t^3e^{r_1t}$，等等。

其次，假设其中的两个根为复根，比如

$$r_5,\ r_6 = h \pm vi$$

则上述解中的第五、六项应合并成如下表达式：

$$e^{ht}(A_5\cos vt + A_6\sin vt) \qquad [\text{参阅 } (15.24')]$$

同理，若求得两个不同的复根，则必然存在两个（具有不同的 $h$ 和 $v$ 值，且二者均有两个任意常数）三角表达式。[①] 作为进一步的可能性，如果恰好存在两对重复根，则我们应采用 $e^{ht}$ 作为其中一对根的乘积项，采用 $te^{ht}$ 作为另一对根的乘积项。而且，即使 $h$ 和 $v$ 在重复根中具有相等的值，也应给每一对复根以不同的一对任意常数。

一旦求得 $y_p$ 和 $y_c$，完备方程(15.50)便可以轻松求出。同前面一样，它不过是余函数与特别积分的和：$y(t) = y_p + y_c$。在此通解中，我们可以算出共有 $n$ 个任意常数。因此，要得到定解，需要 $n$ 个初始条件。

**例1** 求

$$y^{(4)}(t) + 6y'''(t) + 14y''(t) + 16y'(t) + 8y = 24$$

的通解。此四阶方程的特别积分为

546

$$y_p = \frac{24}{8} = 3$$

由(15.51′)，其特征方程为

$$r^4 + 6r^3 + 14r^2 + 16r + 8 = 0$$

---

① 注意到这一点是重要的：由于复根总是以共轭对形式出现，所以我们可以确定，当微分方程为奇数阶，即当 $n$ 为一个奇数时，我们至少可以得到一个实根。

它可分解成如下形式

$$(r+2)(r+2)(r^2+2r+2)=0$$

由前两个加括号的表达式,我们可以得到两个根 $r_1=r_2=-2$,而最后一个二次表达式产生了一对复根 $r_3$, $r_4=-1\pm i$,有 $h=-1$, $v=1$。因而余函数为

$$y_c=A_1e^{-2t}+A_2te^{-2t}+e^{-t}(A_3\cos t+A_4\sin t)$$

通解为

$$y(t)=A_1e^{-2t}+A_2te^{-2t}+e^{-t}(A_3\cos t+A_4\sin t)+3$$

当然,若我们有四个初始条件,可以确定四个常数 $A_1, A_2, A_3$ 和 $A_4$。

注意,本例中所有的特征根或者为负的实根,或者为复根,但其实数部分为负。因此,时间路径必然为收敛的,且瞬时均衡为动态稳定的。

## 收敛性与罗斯定理

解高次特征方程并非总是一个轻松的任务。因此,若我们能找到一种毋需解得特征根便可确定时间路径的敛散性的办法,将具有巨大的益处。幸运的是,确实存在这样一种方法,这种方法可对微分方程进行定性(但不是图形)的分析。

这种方法见诸于罗斯定理。① 表述如下:

当且仅当如下行列式序列

---

① 关于此定理的讨论和其简单证明,参见保罗·萨缪尔森:《经济分析基础》,哈佛大学出版社,1947 年出版,第 429—435 页。

$$|a_1|;\quad \begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix};\quad \begin{vmatrix} a_1 & a_3 & a_5 \\ a_0 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{vmatrix};\quad \begin{vmatrix} a_1 & a_3 & a_5 & a_7 \\ a_0 & a_2 & a_4 & a_6 \\ 0 & a_1 & a_3 & a_5 \\ 0 & a_0 & a_2 & a_4 \end{vmatrix};\cdots$$

的前 $n$ 个行列式都为正时，$n$ 次多项式方程

$$a_0 r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n = 0$$

的所有根的实数部分为负。

在应用此定理时，应记住 $|a_1| \equiv a_1$。而且应理解，对于 $m >$ 547 $n$，我们应取 $a_m = 0$。例如，给定一个三次多项式方程($n = 3$)，我们需要考察上述行列式中前三个行列式的符号；为此，我们应令 $a_4 = a_5 = 0$。

当我们回顾一下，要使时间路径 $y(t)$ 收敛，无论初始条件恰好为什么，微分方程的特征根都必须有负的实数部分，那么，便应知道，此定理与收敛性的相关性是不言自明的。因为特征方程(15.51′)是一个 $n$ 次多项式方程，并有 $a_0 = 1$，所以罗斯定理对检验收敛性具有直接的帮助。事实上，我们注意到特征方程(15.51′)的系数与给定微分方程(15.51)的系数完全一致，所以，假设我们总是取 $a_0 = 1$，则将(15.51)的系数直接代入前面所示的用于检验的行列式序列，是完全可以接受的。由于定理所列的条件是在“当且仅当”的基础上给出的，所以，这些条件显然构成了充要条件。

**例 2** 以罗斯定理检验例 1 中的微分方程是否具有收敛的时间路径。此方程为四次方程，从而 $n = 4$。系数为 $a_0 = 1, a_1 = 6$, $a_2 = 14, a_3 = 16, a_4 = 8, a_5 = a_6 = a_7 = 0$。将其代入前四个行列

式,可求得其值分别为 6、68、800 和 6400。由于它们均为正,所以,我们可以断定,时间路径是收敛的。

## 练习 15.7

---

1　求下面每一方程的特别积分:

(a) $y'''(t)+2y''(t)+y'(t)+2y=8$

(b) $y'''(t)+y''(t)+3y'(t)=1$

(c) $3y'''(t)+9y''(t)=1$

(d) $y^{(4)}(t)+y''(t)=4$

2　求下列方程的 $y_p$ 和 $y_c$(以及通解):

(a) $y'''(t)-2y''(t)-y'(t)+2y=4$

[*Hint*: $r^3-2r^2-r+2=(r-1)(r+1)(r-2)$]

(b) $y'''(t)+7y''(t)+15y'(t)+9y=0$

[*Hint*: $r^3+7r^2+15r+9=(r+1)(r^2+6r+9)$]

(c) $y'''(t)+6y''(t)+10y'(t)+8y=8$

[*Hint*: $r^3+6r^2+10r+8=(r+4)(r^2+2r+2)$]

3　在上题所求得的特征根符号的基础上,分析均衡的动态稳定性。然后再运用罗斯定理检验你的答案。

548　4　毋需求出下列微分方程的特征根,确定它们是否产生收敛的时间路径?

(a) $y'''(t)-10y''(t)+27y'(t)-18y=3$

(b) $y'''(t)+11y''(t)+34y'(t)+24y=5$

(c) $y'''(t)+4y''(t)+5y'(t)-2y=-2$

5　由罗斯定理推导出,对于二阶线性微分方程 $y''(t)+a_1y'(t)+a_2y=b$,无论初始条件如何,当且仅当系数 $a_1$ 和 $a_2$ 均为正时,解的路径才是收敛的。

---

# 第 16 章　离散时间：一阶差分方程

在连续时间情况下，变量 $y$ 的变化模式体现于 $y'(t)$，$y''(t)$ 549 等导数中。所涉及的时间变化在数量上是极微小的。然而，如果将时间视为**离散**变量，因而变量 $t$ 仅取整数值，则导数的概念显然不再适用了。因此，正如我们将要看到的那样，变量 $y$ 的变化模式必须通过 $y(t)$ 的“差分”，而非其导数或微分的概念来描述。相应地，微分方程的方法也需让位于**差分方程**方法。

在涉及离散时间情况下，仅当变量 $t$ 由一个整数值变为另一个整数值时，比如由 $t=1$ 变为 $t=2$ 时，变量 $y$ 的值才会变化。此时，对 $y$ 不做什么假设。在这一点，将 $t$ 值解释成**时期**，而非**时点**（$t=1$ 表示第 1 期，$t=2$ 表示第 2 期如此等等）是很方便的。这样，我们便可以仅将 $y$ 视为在每一时期具有唯一值的量。基于这样的解释，离散时间的动态经济学常被称作**期分析**。但应强调的是，这里使用的“时期”并非日历意义上的时期，而是分析意义上的时期。因此，一个时期在一个具体模型中可能包括一段日历时间，但在另一个模型中则可能包含完全不同的日历时间，而且即使在同一模型中，每一连续时期也不必解释成相等的日历时间。在分 550 析意义上，一个时期仅意味着在变量 $y$ 变化前所逝去的一段时间。

## 16.1 离散时间、差分与差分方程

尽管问题的形成必然会变化，但由连续时间到离散时间的改变，对动态分析的基本性质并无影响。我们的动态问题基本上仍是由已知的变量 $y$ 随时间变化的模式，求出时间路径。但现在的变化模式应以差商 $\Delta y/\Delta t$ 来表示；$\Delta y/\Delta t$ 是导数 $dy/dt$ 在离散时间情况下的对应物。但回忆一下可知，$t$ 现在仅取整数值，这样，当我们将两个连续时期的 $y$ 值相比较时，必然有 $\Delta t = 1$。因此，差商 $\Delta y/\Delta t$ 可以简化为 $\Delta y$，可称其为 *$y$ 的一阶差分*。符号 $\Delta$ 表示差，可以作为取$(y)$的一阶差分的指示。因此，它成为算符 $d/dt$ 在离散时间情况下的对应物。

当然，表达式 $\Delta y$ 可取不同的值，具体取值视取差分(或差分)时所涉及的连续时期而定。为免于模糊，我们给 $y$ 加上时间下标，并将一阶差分更具体地定义如下

$$\Delta y_t \equiv y_{t+1} - y_t \tag{16.1}$$

其中 $y_t$ 表示 $t$ 期 $y$ 值，$y_{t+1}$是紧接着 $t$ 期的那个时期的 $y$ 值。运用这些符号，我们可以将 $y$ 的变化模式以诸如这样的方程

$$\Delta y_t = 2 \tag{16.2}$$

或

$$\Delta y_t = -0.1\, y_t \tag{16.3}$$

来表示。这类方程称作*差分方程*。读者一方面应注意上面两个方程的相似之处，另一方面要注意上面两个方程与微分方程 $dy/dt = 2$ 及 $dy/dt = -0.1t$ 之间的相似性。

尽管差分方程这个名称源于像 $\Delta y_t$ 这样的差式，但仍存在不

含 Δ 表达式且更便于使用的其它方程形式。根据 16.1,我们可将(16.2)改写成

$$y_{t+1}-y_t=2 \tag{16.2'}$$

或

$$y_{t+1}=y_t+2 \tag{16.2''}$$

至于(16.3),相应的等价形式为

551

$$y_{t+1}-0.9y_t=0 \tag{16.3'}$$

或

$$y_{t+1}=0.9y_t \tag{16.3''}$$

当由已知的上一期的 $y$ 值计算本期 $y$ 值时,(16.2″)和(16.3″)那种形式是方便的。但在后面的讨论中,我们将更经常地使用(16.2′)和(16.3′)那种形式。

注意到在差分方程中,时间下标的选择多少有些任意的是重要的。例如,将(16.2′)改写成 $y_t-y_{t-1}=2$,其中$(t-1)$表示紧接着 $t$ 期的那个时期,这丝毫不改变方程的意义。或者,我们可以将其等价地表示成 $y_{t+2}-y_{t+1}=2$。

还应指出的是,尽管我们一致地使用带下标的 $y$ 符号,但使用符号 $y(t)$,$y(t+1)$及 $y(t-1)$等,也是可以的。但是为避免在连续时间和离散时间两种情况下均使用符号 $y(t)$,在期分析的讨论中,我们将坚持使用带下标的符号。

与微分方程类似,差分方程可以是线性的,也可以是非线性的;可以是齐次的,也可以是非齐次的;既可以是一阶的,也可以是二阶或更高阶的。以(16.2′)为例,可以将其分类如下:(1)它是线性的,因为没有任意一期的 $y$ 项自乘至二次或更高次幂,也没有 $y$

项与另一期的 $y$ 项相乘。(2)它是非齐次的,因为方程右边(其中不含 $y$ 项)为非零。(3) 它是一阶的,因为仅存在一阶差分 $\Delta y_t$,仅含有一期时滞。(将在 17 章中讨论的二阶差分方程则与此相反,它包含两期时滞,因而含有三个 $y$ 项:$y_{t+2}$,$y_{t+1}$及 $y_t$。)

实际上,(16.2′)也具有含常系数和常数项的特征。因为常系数是我们要考察的唯一情况,所以这个特征是毫无疑问的。常数项这一特征也将贯穿本章始终,而处理可变项情况的方法则在 17 章中讨论。

## 16.2 解一阶差分方程

在解微分方程时,我们的目标是求出时间路径 $y(t)$。正如我们所知道的那样,此时间路径是时间的函数,它不含任何导数或微分表达式,并与给定微分方程及其初始条件完全一致。我们由差分方程所求出的时间路径也具有类似的性质。它也应是 $t$ 的函数——一个在每个时期中定义 $y$ 值的公式——它与给定的差分方程及其初始条件是一致的。此外,它一定不含诸如 $\Delta y_t$(或像 $y_{t+1}-y_t$)等的任何差分表达式。

552

在最终分析中,解微分方程是一个积分问题。那么,解差分方程呢?

### 迭代法

在提出解差分方程的一般方法之前,我们首先介绍一种相对平淡的方法,迭代法;尽管这种方法相对粗糙,但能深刻地揭示所谓"解"的实质。

在本章，我们仅涉及一阶差分方程的情况。因此，差分方程仅描述两个连续时期间的 $y$ 的变化模式。一旦这样的变化模式给定，如以(16.2″)那种方式给定，一旦初始值 $y_0$ 给定，则由方程求出 $y_1$ 便没有问题。类似地，一旦求得 $y_1$，则通过重复应用(迭代)差分方程所设定的变化模式，马上就可以得到 $y_2$，等等。这样，迭代的结果便使我们得以导出时间路径。

**例 1** 求差分方程(16.2)的解，假定初始值 $y_1 = 15$。为完成迭代过程，运用(16.2″)这种形式的差分方程，即 $y_{t+1} = y_t + 2$，$y_0 = 15$ 这种形式，要更方便一些。由此方程，我们可一步步导出

$$y_1 = y_0 + 2$$

$$y_2 = y_1 + 2 = (y_0 + 2) + 2 = y_0 + 2(2)$$

$$Y_3 = y_2 + 2 = [y_0 + 2(2)] + 2 = y_0 + 3(2)$$

.........................................

一般而言，对任意时期 $t$

$$y_t = y_0 + t(2) = 15 + 2t \tag{16.4}$$

上面的方程给出了任意时期的 $y$ 值(包括起始时期 $t = 0$ 的 $y$ 值)；因此，它构成了(16.2)的解。

迭代过程是粗糙的，它对应于以直接积分粗糙地解简单微分方程的方法，但它可以明确指出产生时间路径的方式。一般而言，$y_t$ 值以设定的方式依赖于紧邻的上一期的 $y$ 值($y_{t-1}$)。因此，一个已知的初始值 $y_0$ 可通过所描述的变化模式，依次导致 $y_1, y_2$，……。

**例 2** 解差分方程(16.3)。这次，不设定初始值，仅以 $y_0$ 表示之。同样，处理另一种形式(16.3″)即 $y_{t+1} = 0.9y_t$ 要更方便一

553 些。通过迭代，我们有

$$y_1 = 0.9y_0$$

$$y_2 = 0.9y_1 = 0.9(0.9y_0) = (0.9)^2 y_0$$

$$y_3 = 0.9y_2 = 0.9(0.9)^2 y_0 = (0.9)^3 y_0$$

……………………………………

可将其概括成解

$$y_t = (0.9)^t y_0 \tag{16.5}$$

为提高读者的兴趣，我们可以给此例赋予一些经济意义。在简单的乘数分析中，第 0 期的一次投资会导致诸轮支出，这依次又会在随后各期产生不同的收入增量。用 $y$ 表示**收入增量**，我们有 $y_0$ = 第 0 期的投资量；但下一期的收入增量将取决于边际消费倾向(MPC)。若 MPC = 0.9，且若每期收入仅在下一时期消费，则 90%的 $y_0$ 将在第 1 期消费，导致时期 1 的收入增量 $y_1 = 0.9y_0$。通过类似推理，可求得 $y_2 = 0.9y_1$，等等。我们可以看到，这恰好是上述迭代过程的结果。换言之，产生收入的乘数过程可以通过像(16.3″)这样的差分方程来描述，而像(16.5)这样的解则可告诉我们在任意时期 $t$，收入增量的大小。

**例 3** 解齐次差分方程

$$my_{t+1} - ny_t = 0$$

通过正规化并移项，此方程可以写成

$$y_{t+1} = \left(\frac{n}{m}\right) y_t$$

除了以 $n/m$ 代替了 0.9 以外，此方程与例 2(16.3″)基本相同。因此，通过类比可知，解应为

$$y_t = (\frac{n}{m})^t y_0$$

考察$(\frac{n}{m})^t$项。正是通过这项，不同的 $t$ 值将产生其对应的 $y$ 值。因此，它对应于微分方程解中的表达式 $e^{rt}$。若我们更一般地将其写成 $b^t$($b$ 为底)并加上更一般的乘积常数 $A$(而非 $y_0$)，我们可以看到一般齐次差分方程例 3 的解的形式为

$$y_t = Ab^t$$

我们将发现，$Ab^t$ 式在差分方程中所起的作用与 $Ae^{rt}$ 式在微分方程中所起的作用同样重要。[①] 但是，尽管二者均为指数表达式，但前者的底为 $b$，后者的底为 $e$。由此可以推断，正如连续时间路径 $y(t)$ 的类型极大地依赖于 $r$ 值一样，离散时间路径 $y_t$ 主要依赖于 $b$ 值。

554

## 一般方法

读者现在对微分方程和差分方程间的各种相似性必然留有深刻的印象。可以推测，现在将要解释的差分方程的一般解法与微分方程的一般解法也将有很多的相似之处。

假设我们正在试图求解一阶差分方程

$$y_{t+1} + ay_t = c \tag{16.6}$$

其中 $a$ 和 $c$ 为两个常数。其通解将由两部分的和构成：特别积分

① 读者可能会指出例 1 的解(16.4)并不包含形式为 $Ab^t$ 项，因而反对这一观点。但事实上，这一结果的产生只是由于在例 1 中，我们有 $b = n/m = 1/1 = 1$，因而 $Ab^t$ 项简化为常数。

$y_p$,[①] [它是完备非齐次方程(16.6)的任意解,]以及余函数 $y_c$[它是(16.6)的简化方程的通解]。(16.6)的简化方程为:

$$y_{t+1} + ay_t = 0 \tag{16.7}$$

$y_p$ 部分仍表示 $y$ 的瞬时均衡水平,$y_c$ 部分表示时间路径与均衡的偏差。$y_p$ 与 $y_c$ 的和构成了差分方程的通解,是由于存在一个任意常数。同以前一样,要得到定解,则需知道初始条件。

我们首先讨论余函数。解例 3 的经验表明,我们可以试探形式为 $y_t = Ab^t$($Ab^t \neq 0$,否则 $y_t$ 会成为位于 $t$ 轴上的一条水平线)的解;在此情况下,我们也有 $y_{t+1} = Ab^{t+1}$。如果 $y_t$ 和 $y_{t+1}$这些值成立,则齐次方程(16.7)变成

$$Ab^{t+1} + aAb^t = 0$$

削去非零的公共因子 $Ab^t$,产生

$$b + a = 0 \text{ 或 } b = -a$$

这意味着要使试探解成立,我们必须令 $b = -a$。因此余函数可以写成

$$y_c (= Ab^t) = A(-a)^t$$

现在我们来求与完备方程(16.6)相联系的特别积分。对这个问题,例 3 没有任何帮助,因为它仅与齐次方程有关。但我们要注意,对于 $y_p$,我们可以选择(16.6)的任何解;因此,若形式最为简

555 单的试探解 $y_t = k$(常数)成立,则我们便不会遇到实质的困难。现在,若 $y_t = k$,则 $y$ 会在不同时间保持同样的常数值,因此我们必然有 $y_{t+1} = k$。将这些值代入(16.6)得到

---

① 尽管这里并不涉及"积分",但我们仍从微分方程中借用这一术语。也有一些作者称其为特别解。

$$k + ak = c \quad 和 \quad k = \frac{c}{1+a}$$

因为此特定的 $k$ 值满足方程,所以特别积分可以写成

$$y_p(=k) = \frac{c}{1+a} \quad (a \neq -1)$$

这是一个常数,所以在这种情况下它是一个稳定均衡。

但是,若像例 1 中那样,恰好 $a=-1$,那么,特别积分 $c/(1+a)$没有定义,这样必须求非齐次方程(16.6)的其它解。在这种情况下,我们采用现在已经熟悉了的形式为 $y_t = kt$ 的试探解。当然,这意味着 $y_{t+1} = k(t+1)$。将其代入(16.6),求得

$$k(t+1) + akt = c \quad 和 \quad k = \frac{c}{t+1+at} = c$$

$$[因为\ a=-1] \quad 因此\ y_p(=kt) = ct$$

这种形式的特别积分是 $t$ 的非常数函数,因此它表示移动均衡。

将 $y_c$ 与 $y_p$ 加在一起,我们现在可将通解写成如下两种形式中的一种:

$$y_t = A(-a)^t + \frac{c}{1+a} \quad [通解, a \neq -1 的情况] \tag{16.8}$$

$$y_t = A(-a)^t + ct = A + ct \quad [通解, a = -1 的情况] \tag{16.9}$$

由于存在任意常数 $A$,这两个解没有一个是完全确定的。为消除任意常数,我们借助于初始条件当 $t=0$ 时,$y_t = y_0$,令(16.8)中的 $t=0$,我们有

$$y_0 = A + \frac{c}{1+a} \quad 和 \quad A = y_0 - \frac{c}{1+a}$$

因而,(16.8)的确定形式为

$$y_t = \left(y_0 - \frac{c}{1+a}\right)(-a)^t + \frac{c}{1+a} \quad [\text{定解}, a \neq -1 \text{ 的情况}]$$

另一方面，令(16.9)中的 $t=0$，求得 $y_0 = A$，从而(16.9)的确定形式为

$$y_t = y_0 + at \quad [\text{定解}, a = -1 \text{ 的情况}] \qquad (16.9')$$

556若将最后一个结果应用于前面例1，那么，所得到的解与迭代解(16.4)是完全相同的。

读者可以通过如下两个步骤，检验上述每个解的正确性。首先，令(16.8′)中的 $t=0$，可以看到此方程化简为等式 $y_0 = y_0$，表明它满足初始条件。其次，将 $y_t$ 公式(16.8′)和类似的 $y_{t+1}$ 公式(通过将(16.8′)中的 $t$ 置换成 $t+1$，便可得到)代入(16.6)，可以看到(16.6)化简为等式 $c=c$，表明时间路径与给定差分方程是一致的。解(16.9′)正确性的检验，也是类似的。

**例4** 解一阶差分方程

$$y_{t+1} - 5y_t = 1 \quad \left(y_0 = \frac{7}{4}\right)$$

根据推导(16.8′)的步骤，我们可通过试探解 $y_t = Ab^t$(它意味着 $y_{t+1} = Ab^{t+1}$，)可求得 $y_c$。将这些值代入齐次形式 $y_{t+1} - 5y_t = 0$，并取消公共因子 $Ab^t$，得到 $b=5$。因此，

$$y_c = A(5)^t$$

为求得 $y_p$，试探解 $y_t = k$，这意味着 $y_{t+1} = k$。将其代入完备差分方程，求得 $k = -\frac{1}{4}$。因此

$$y_p = -\frac{1}{4}$$

由此可知，通解为

$$y_t = y_c + y_p = A(5)^t - \frac{1}{4}$$

这里令 $t=0$ 并运用初始条件 $y_0 = \frac{7}{4}$，得到 $A=2$。因此，最终定解可以写成

$$y_t = 2(5)^t - \frac{1}{4}$$

因为本例给出的差分方程是(16.6)在 $a=-5, c=1$ 及 $y_0 = 7/4$ 时的特例，又因为(16.8′)是这类差分方程的解的“公式”，所以，我们也可以在(16.8′)中加入具体参数值而求得我们要求的解，结果为

$$y_t = \left(\frac{7}{4} - \frac{1}{1-5}\right)(5)^t + \frac{1}{1-5} = 2(5)^t - \frac{1}{4}$$

它与前面的答案完全一致。

注意，(16.6)中的 $y_{t+1}$ 项具有单位系数。若已知差分方程中的这项不具有单位系数，在运用解公式(16.8′)前必须将其正规化。

## 练习 16.2

---

1 将下列差分方程变换为(16.2″)的形式：

(a) $\Delta y_t = 7$ (b) $\Delta y_t = 0.3y_t$ (c) $\Delta y_t = 2y_t - 9$

557

2 用迭代法解下列差分方程

(a) $y_{t+1} = y_t + 1$ $(y_0 = 10)$

(b) $y_{t+1} = \alpha y_t$ $(y_0 = \beta)$

(c) $y_{t+1} = \alpha y_t - \beta$ ($y_t = y_0$ 当 $t=0$)

3 按(16.6)式改写上题中的差分方程，并用公式(16.8′)或(16.9′)解之。(哪个方便，便用哪个。)答案与用迭代法求得的答案一致吗？

4 对于下列每个差分方程，运用推导(16.8′)和(16.9′)所描述的步骤，

求 $y_c$ 和 $y_p$，以及定解：

(a) $y_{t+1}+3y_t=4$ $(y_0=4)$

(b) $2y_{t+1}-y_t=6$ $(y_0=7)$

(c) $y_{t+1}=0.2y_t+4$ $(y_0=4)$

---

## 16.3 均衡的动态稳定性

在连续时间情况下，均衡的动态稳定性取决于余函数中的 $Ae^{rt}$ 项。在期分析中，余函数中的 $Ab^t$ 则起着同样的作用。因为 $Ab^t$ 的解释要比 $Ae^t$ 稍微复杂一些，所以我们在进行下一步之前首先澄清其含义。

### b 的意义

均衡是否是动态稳定的问题也就是当 $t\to\infty$ 时余函数是否趋于零的问题。从根本上看，我们必须分析当 $t$ 无限增大时 $Ab^t$ 项的路径。显然，$b$ 值(指数项的底)在这方面具有关键作用。我们首先抛开系数 $A$(假设 $A=1$)，单独考察 $b$ 的含义。

为分析之目的，我们首先将 $b$ 的可能的值域分成七个不同的区域，按着 $b$ 值的递减顺序，列在表 16.1 的前两列中。这些区域也标在图 16.1 纵轴 $b$ 的刻度上，并以 $+1$，$0$ 和 $-1$ 作为分界点。事实上，这三个点本身构成了区域 II、IV 和 VI。而区域 III 和 V 分别对应所有正分数和所有负分数的集合。余下的两个区域 I 和 VII，则是 $b$ 的绝对值超过 1 的区域。

558 在每个不同的区域，指数式 $b^t$ 产生了不同的时间路径。在表

表 16.1　$b$ 值的分类

| 区域 | $b$ 值 | | $b^t$ 的值 | 不同时期的 $b^t$ 值 | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | $t=0$ | $t=1$ | $t=2$ | $t=3$ | $t=4\cdots$ |
| I | $b>1$ | $(\|b\|>1)$ | 例如$(2)^t$ | 1 | 2 | 4 | 8 | 16 |
| II | $b=1$ | $(\|b\|=1)$ | $(1)^t$ | 1 | 1 | 1 | 1 | 1 |
| III | $o<b<1$ | $(\|b\|<1)$ | 例如$(\frac{1}{2})^t$ | 1 | $\frac{1}{2}$ | $\frac{1}{4}$ | $\frac{1}{8}$ | $\frac{1}{16}$ |
| IV | $b=0$ | $(\|b\|=0)$ | $(0)^t$ | 0 | 0 | 0 | 0 | 0 |
| V | $-1<b<0$ | $(\|b\|<1)$ | 例如$(-\frac{1}{2})^t$ | 1 | $-\frac{1}{2}$ | $\frac{1}{4}$ | $-\frac{1}{8}$ | $\frac{1}{16}$ |
| VI | $b=-1$ | $(\|b\|=1)$ | $(-1)^t$ | 1 | $-1$ | 1 | $-1$ | 1 |
| VII | $b<-1$ | $(\|b\|>1)$ | 例如$(-2)^t$ | 1 | $-2$ | 4 | $-8$ | 16 |

16.1 和图 16.1 中已分别举例说明。在区域 I(其中 $b>1$),$b^t$ 必然随 $t$ 以递增的速度增加。因此时间路径的图形绘成图 16.1 最上面的图形。注意,这个图形绘成阶梯函数而非平滑曲线,这是因为我们研究的是期分析问题。在区域 II($b=1$)中,对所有的 $t$ 值,$b^t$ 均为 1。因此其图形为一条水平直线。接下来,在区域 III 中,$b^t$ 表示正分数自乘至整数幂。当幂数增加时,尽管 $b^t$ 仍然为正,但必然递减。在区域 IV,$b=0$ 的情况与 $b=1$ 的情况极为类似,但这里我们有 $b^t=0$ 而非 $b^t=1$,因而其图形与横轴重合。但这种情况仅具有表面意义,因为前面我们已采纳了 $Ab^t\neq0$ 的假设,这意味着 $b\neq0$。

559

当我们移至负的区域时,产生了一种有趣的现象:从一个时期

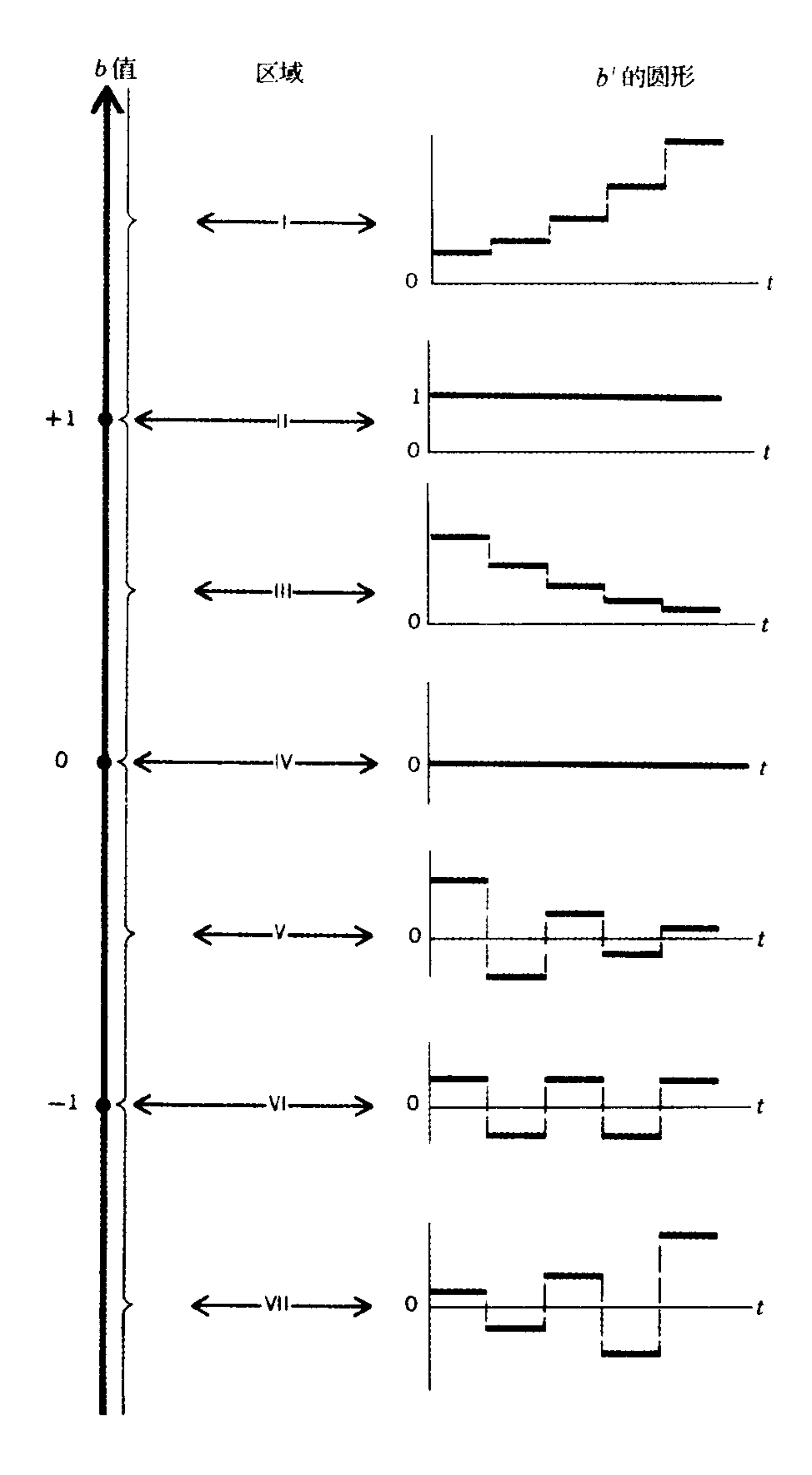



图 16.1

到下个时期，$b^t$ 的值交替变换正负号。这一事实清楚地在表 16.1 的后三行和图 16.1 中下面的三个图形中给出。在区域 V，$b$ 为负分数，交替变换的时间路径越来越趋近于横轴（参见区域 III 中的正分数）。相反，当 $b=-1$（区域 VI）时，则会产生在 $+1$ 和 $-1$ 间持久变换的时间路径。最后，当 $b<-1$ 时，（区域 VII），交替变换的时间路径会越来越远离横轴。

引人注意的是，时间路径波动的现象虽然不可能仅由 $Ae^{rt}$ 项产生（二阶微分方程的复根情况要求一对复根），但是，波动却可能仅由单个的 $b^t$（或 $Ab^t$）产生。但要注意，波动的特征却有些不同；有别于三角函数的波动模式，图 16.1 所描述的波动是非平滑的。由于这个原因，我们将使用**振荡**一词来表示这种新的非平滑的波动形式，尽管很多作者确实不加区别地使用“波动”与“振荡”这两个词。

上面讨论的实质可通过如下一般陈述来表达： 560

$$若\begin{cases}b>0\\b<0\end{cases}，则\ b^t\ 的时间路径将是\begin{cases}非振荡的\\振荡的\end{cases}$$

$$若\begin{cases}|b|>1\\|b|<1\end{cases}，则\ b^t\ 的时间路径将是\begin{cases}发散的\\收敛的\end{cases}$$

注意到这一点是重要的：虽然 $e^{rt}$ 式的收敛性取决于 $r$ 的符号，但 $b^t$ 式的收敛性却依赖于 $b$ 的**绝对值**。

### A 的作用

迄今为止，我们一直谨慎地将乘积常数 $A$ 排除在外。但其影响（共有两方面）是相对易于考虑的。首先，$A$ 的**大小**用于“放大”（比如 $A=3$）或“收缩”（比如若 $A=1/5$）$b^t$ 的值。即它不改变时

间路径的基本图形，但能产生标度效应。而 $A$ 的符号确实从根本上影响时间路径的图形，因为若 $b^t$ 为 $A=-1$ 相乘，则图 16.1 所示的时间路径将为其自身对横轴的镜像所置换。因此，负 $A$ 可以产生镜像效应及标度效应。

## 收敛于均衡

上述讨论对余函数中的 $Ab^t$ 项进行了解释，即它表示与某均衡水平的偏差。若在 $Ab^t$ 项上加上一项，比如 $y_p=5$，则时间路径必然垂直上移常数 5。这丝毫不影响时间路径的敛散性，但却改变了测定收敛或发散时所参照的水平。图 16.1 所描述的是 $Ab^t$ 式收敛（或不收敛）于零。但当包含 $y_p$ 时，便变成了时间路径 $y_t=y_c+y_p$ 收敛于均衡水平 $y_p$ 的问题。

关于这一问题，我们再对 $b=1$ 的特殊情况（区域 II）多做一点解释。下面的时间路径

$$y_t = A(1)^t + y_p = A + y_p$$

给人一种收敛的印象，因为乘积项 $(1)^t=1$ 不产生任何发散的效果。但观察一下，$y_t$ 现在取值 $(A+y_p)$，而非均衡值 $y_p$；实际上，它永远也不可能达到 $y_p$，除非 $A=0$。作为对这种情况的一种解释，我们引用（16.9）中的一个时间路径，它涉及一个移动均衡 $y_p=ct$。这个时间路径是一个发散的路径，但原因不在于在特别积分中出现 $t$，而在于非零的 $A$，与移动均衡存在一个固定的偏差。因此，在规定时间路径 $y_t$ 对均衡 $y_p$ 的收敛条件时，必须排除 $b=1$ 的情况。

总之，当且仅当 $|b|<1$ 时，解

$$y_t = Ab^t + y_p$$

是一个收敛的路径。

**例 1**　$y_t = 2(-\frac{4}{5})^t + 9$ 表示哪类时间路径？因为 $b = -4/5 < 0$，所以时间路径是振荡的。但是因为 $|b| = \frac{4}{5} < 1$，所以振荡是衰减的，且时间路径收敛于均衡水平 9。

读者应小心区别 $2(-4/5)^t$ 和 $-2(\frac{4}{5})^t$。它们表示完全不同的时间路径图形。

**例 2**　如何描述时间路径 $y_t = 3(2)^t + 4$ 的特征？因为 $b = 2 > 0$，所以不会产生振荡。但因 $|b| = 2 > 1$，所以时间路径将发散于均衡水平 4。

## 练习 16.3

---

1　讨论下列时间路径的性质：

(a) $y_t = 3^t + 1$　　(c) $y_t = 5(-\frac{1}{10})^t + 3$

(b) $y_t = 2(\frac{1}{3})^t$　　(d) $y_t = -3(\frac{1}{4})^t + 2$

2　在练习 16.2－4 中求得的每个差分方程的时间路径具有何种性质？

3　求下列方程的解，并确定时间路径是否是振荡且收敛的：

(a) $y_{t+1} - \frac{1}{3}y_t = 6$　　$(y_0 = 1)$

(b) $y_{t+1} + 2y_t = 9$　　$(y_0 = 4)$

(c) $y_{t+1} + \frac{1}{4}y_t = 5$　　$(y_0 = 2)$

(d) $y_{t+1} - y_t = 3$　　$(y_0 = 5)$

---

## 16.4 蛛网模型

为介绍一阶差分方程在经济分析中的应用，我们将引证两个单一商品市场模型的变形。第一个变形称作**蛛网模型**，它与我们前面介绍的市场模型的不同之处在于：它不将 $Q_s$ 作为现期价格的函数，而作为前一期价格的函数。

### 模 型

考察这样一种情境：生产者的产出决策必须在实际销售之前562做出——比如农业生产，种植必须比收获及产出的销售早一个适当的时期。我们假设 $t$ 期的产出决策是基于当时流行价格 $P_t$。但因产出直到 $(t+1)$ 期方能销售，所以，$P_t$ 不能确定 $Q_{st}$，只能确定 $Q_{s,t+1}$。因此，我们现在有了一个"滞后"供给函数[①]

$$Q_{s,t+1} = S(P_t)$$

或者等价地，后移一期时间下标

$$Q_{st} = S(P_{t-1})$$

当这个供给函数与形式为

$$Q_{dt} = D(P_t)$$

的需求函数相互作用时，使会产生一种有趣的动态价格模式。

取线性形式的(滞后)供给函数和(非滞后)的需求函数，并假

① 我们这里作一个隐含的假设：一个时期的产出将全部投放市场，不留存货。当所研究商品属易坏商品或未曾有存货的商品时，作此假设是合适的。含有存货的模型将在下节考察。

定每一时期的市场价格均处于市场出清时的价格水平，则我们有含如下三个方程的市场模型

$$Q_{dt} = Q_{st}$$
$$Q_{dt} = \alpha - \beta P_t \qquad (\alpha, \beta > 0)$$
$$Q_{st} = -\gamma + \delta P_{t-1} \qquad (\gamma, \delta > 0) \qquad (16.10)$$

但是，将后两个方程代入第一个方程，模型可以化为一个一阶差分方程

$$\beta P_t + \delta P_{t-1} = \alpha + \gamma$$

为解此方程，应首先将其正规化，并将时间下标向前移一期[即变 $t$ 为$(t+1)$等]。结果有

$$P_{t+1} + \frac{\delta}{\beta} P_t = \frac{\alpha + \gamma}{\beta} \qquad (16.11)$$

它是作了如下代换的(16.6)的复制品：

$$y = P \quad a = \frac{\delta}{\beta} \quad 和 \quad c = \frac{\alpha + \gamma}{\beta}$$

由于$\delta$ 和 $\beta$ 均为正，所以 $a \neq -1$。因而，我们可以应用公式(16.8′)，得到时间路径

$$P_t = \left(P_0 - \frac{\alpha + \gamma}{\beta + \delta}\right)\left(-\frac{\delta}{\beta}\right)^t + \frac{\alpha + \gamma}{\beta + \delta} \qquad (16.12)$$

其中 $P_0$ 表示初始价格。

## 蛛 网

563

关于此模型，可以观测到以下三点：首先，$(\alpha + \gamma)/(\beta + \delta)$式构成了差分方程的特别积分，可以将其视为模型的瞬时均

衡价格：①

$$\bar{P}=\frac{\alpha+\gamma}{\beta+\delta}$$

因为它是一个常数，所以是稳定均衡。将 $\bar{P}$ 代入解中，可将时间路径 $P_t$ 表示成另一种形式

$$P_t=(P_0-\bar{P})(-\frac{\delta}{\beta})^t+\bar{P} \tag{16.12'}$$

此方程引出了第二点，即表达式$(P_0-\overline{P})$的意义。因为$(P_0-\overline{P})$对应于 $Ab^t$ 项中的常数 A，所以，$(P_0-\overline{P})$的符号决定时间路径是从均衡水平以上开始还是从以下开始（镜像效应），而其大小则决定与均衡水平的远近（标度效应）。最后一点，还有式$(-\delta/\beta)$，它对应于 $Ab^t$ 中的 $b$ 部分。根据模型的设定的 $\beta>0,\delta>0$，我们可导出一个振荡的时间路径。正是这一事实导致了我们即将看到的蛛网现象。当然，此模型可能产生三种可能类型的振荡。根据表 16.1 或图 16.1 可知，

$$\text{若}\delta \begin{matrix} > \\ = \\ < \end{matrix} \beta\text{，振荡将为}\begin{cases}\text{放大振荡}\\ \text{单位振荡}\\ \text{衰减振荡}\end{cases}$$

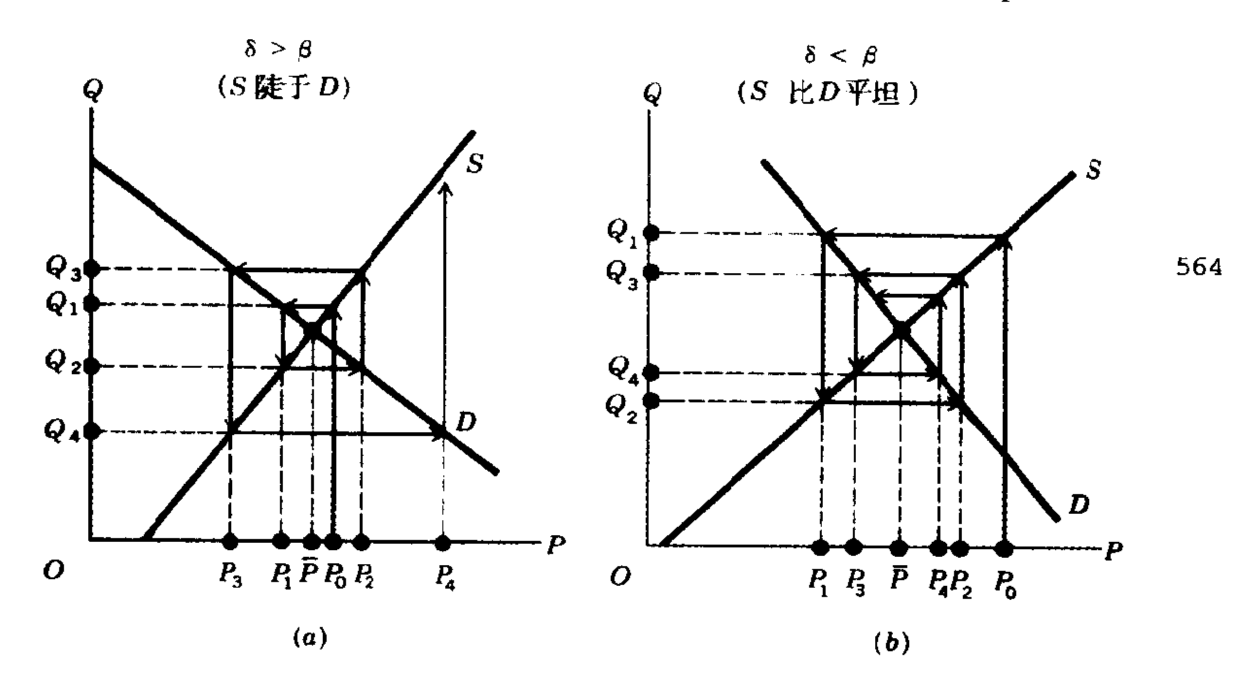
其中“单位振荡”，是指区域 VI 的路径类型。

为使蛛网直观化，我们将模型(16.10)绘在图 16.2 中。(16.10)中的第二个方程绘成了一条向下倾斜的线性需求曲线，其斜率在数量上等于 $\beta$。类似地，斜率为$\delta$ 的线性供给曲线可由第三个方程绘出，当然，这时 $Q$ 轴要表示滞后的供给量。$\delta>\beta$（$S$陡于$D$）

① 就市场出清意义的均衡而言，在每一时期所达到的价格都是均衡价格，因为我们已假设，对于每个 $t$，$Q_{dt}=Q_{st}$。

和 $\delta<\beta$($S$ 比 $D$ 平坦)的情况分别绘在图 $a$ 和图 $b$ 中。但在每种情况下,$D$ 和 $S$ 的交点都将产生瞬时均衡价格 $\bar{P}$。

如图 $a$,当 $\delta>\beta$ 时,供求的相互作用将会产生如下放大振荡。给定初始价格 $P_0$(这里假设高于 $\bar{P}$),顺着箭头,我们可在 $S$ 曲线上读出下一期的供给量(第 1 期)将为 $Q_1$。为使市场出清,第 1 期的需求量必须也为 $Q_1$,而这当且仅当价格确定在 $P_1$ 时,方能




564

图 16.2

做到(见向下的箭头)。现在,根据 $S$ 曲线,价格 $P_1$ 会导致在第 2 期产生 $Q_2$ 的供给量,且为使市场在第 2 期出清,按照需求曲线,价格必须定在 $P_2$ 水平。重复这一推理,顺着图中的箭头,我们便可以依次推出以后各期的价格和数量,围绕着供求曲线结成"蛛网"。比较价格水平 $P_0,P_1,P_2,\cdots\cdots$,我们不仅可以观察到振荡的变化模式,而且也可以观测到,随着时光流逝,价格与均衡的偏离不断扩大的倾向。具有这种由内向外结成的蛛网的时间路径是发散的,振荡是放大的。

与其相对照的是，在图 $b$ 中，$\delta < \beta$，会结成一个指向中心的网。若我们顺着箭头从 $P_0$ 出发，我们将越来越接近于供求曲线的交点，即价格水平为 $\bar{P}$ 的位置。价格路径虽然是振荡的，但却是收敛的。

在图 16.2 中，我们未绘出第三种可能性，即$\delta = \beta$ 时的图形。但它所涉及的图形分析方法与前两种情况是完全类似的。因此，我们把它留给读者作为练习。

上述讨论仅涉及 $P$ 的时间路径。但在求得 $P_t$ 后，稍作努力，便可得到 $Q$ 的时间路径。(16.10)的第二个方程将 $Q_{dt}$ 与 $P_t$ 联系起来，从而，若将(12.12)或(12.12′)代入需求方程，便可马上得到 $Q_{dt}$ 的时间路径。进而，因为在每一时期，$Q_{dt}$ 必定等于 $Q_{st}$（市场出清），所以我们可以将此时间路径仅称作 $Q_t$ 而非 $Q_{dt}$。在图 16.2 中，这种替代的合理性是很容易看出的。$D$ 曲线上的每一个点将 565 $P_i$ 与同期的 $Q_i$ 联系起来，因而需求函数能够将价格的时间路径映射到数量的时间路径上去。

读者应注意，图 16.2 的图形方法即便在 $D$ 曲线和 $S$ 曲线为非线性时，也是可以应用的。

## 练习 16.4

---

1　在(16.10)的基础上，分析 $Q$ 的时间路径，分析其收敛性的条件。

2　绘出类似于图 16.2 的图形，并证明，在$\delta = \beta$ 的情况下，价格路径将是既不衰减也不放大的单位振荡。

3　给定蛛网模型的如下需求和供给函数，求瞬时均衡价格，并确定均衡是否稳定：

(a) $Q_{dt} = 18 - 3P_t$ $\qquad$ $Q_{st} = -3 + 4P_{t-1}$

(b) $Q_{dt} = 22 - 3P_t$ $\qquad$ $Q_{st} = -2 + P_{t-1}$

(c) $Q_{dt} = 19 - 6P_t$ $\qquad$ $Q_{st} = 6P_{t-1} - 5$

4 在模型(16.10)中,令条件 $Q_{dt} = Q_{st}$ 及需求函数不变,但供给函数变成

$$Q_{st} = -\gamma + \delta P_t^*$$

其中 $P_t^*$ 表示 $t$ 期的预期价格。进而,假定卖者具有"适应性"价格预期:①

$$P_t^* = P_{t-1}^* + \eta(P_{t-1} - P_{t-1}^*) \quad (0 < \eta \leqslant 1)$$

其中$\eta$(希腊字母 *eta*)为预期调整系数。

(a) 给出上述方程的经济解释。它与适应性预期方程(15.34)在哪些方面类似,哪些方面不同?

(b) 若$\eta$取其极大值,会出现何结果?我们可将蛛网模型视为现在模型的一个特例吗?

(c) 证明:这个新的模型可通过一阶差分方程

$$P_{t+1} - \left(1 - \eta - \frac{\eta\delta}{\beta}\right)P_t = \frac{\eta(\alpha + \gamma)}{\beta}$$

来表示。[提示:解供给函数求 $P_t^*$,然后再运用 $Q_{st} = Q_{dt} = \alpha - \beta P_t$ 这一信息。]

(d) 求价格的时间路径。此时间路径必然是振荡的吗?它能够振荡吗?在什么条件下能振荡?

(e) 证明:时间路径 $P_t$ 如果是振荡的,则仅当 $1 - 2/\eta < -\delta/\beta$ 时,才会是收敛的。

5 蛛网模型,同前面遇到的动态市场模型一样,实质上是以 3.2 节给出的静态市场模型为基础的。在现在的模型中,何种经济假设是其动态化的关键?请解释之。

---

① 参见麦克·纳洛夫"适应性预期与蛛网现象",载《经济学季刊》,1958 年 5 月,227—240 页。

# 16.5 一个具有存货的市场模型

在前一个模型中,我们假设价格以出清每期产出的方式来确定。该假设的含义是:该商品或者是易坏、不能储存的,或者是虽可储存,但未曾留有存货。现在我们来构建一个卖者保有商品存货的模型。

## 模 型

我们作如下假设:

1 需求量 $Q_{dt}$ 和现期产出量 $Q_{st}$ 均是价格 $P_t$ 的非滞后线性函数。

2 价格调整不是通过每期市场出清来进行的,而是通过卖者的定价过程来实现的:在每期的开始,卖者在考虑到存货状况以后为该期确定一个价格。如果上期价格使存货积累下来,则确定一个比上期低一些的价格;但若存货告罄,则确定一个比上期高的价格。

3 从一期到另一期的价格调整与观测到的存货变化成反比。

根据这些假设,我们可以写出如下方程

$$Q_{dt} = \alpha - \beta P_t \qquad (\alpha, \beta > 0)$$

$$Q_{st} = -\gamma + \delta P_1 \qquad (\gamma, \delta > 0)$$

$$P_{t+1} = P_t - \sigma (Q_{st} - Q_{dt}) \qquad (\sigma > 0) \tag{16.13}$$

其中$\sigma$ 表示存货引致的价格调整系数。注意,(16.13)其实是 14.2 节市场模型在离散时间条件下的对应物,只不过现在的价格调整过程是按存货($Q_{st} - Q_{dt}$)而非按超额需求($Q_{dt} - Q_{st}$)来表示的。

但分析结果已有很大不同;因为在离散时间情况下,我们可能遇到振荡现象。下面,我们来推导并分析时间路径。

### 时间路径

把前两个方程代入第三个方程,模型可以化为一个差分方程:

$$P_{t+1} - [1 - \sigma(\beta + \delta)]P_t = \sigma(\alpha + \gamma) \qquad (16.14)$$

其解由(16.8′)给出:

567

$$\begin{aligned} P_t &= \left(P_0 - \frac{\alpha + \gamma}{\beta + \delta}\right)[1 - \sigma(\beta + \delta)]^t + \frac{\alpha + \gamma}{\beta + \delta} \\ &= (P_0 - \bar{P})[1 - \sigma(\beta + \delta)]^t + \bar{P} \end{aligned} \qquad (16.15)$$

因此,模型的动态稳定性显然取决于 $1 - \sigma(\beta + \delta)$式。为简便计,我们将此式以 $b$ 表示。

参照表 16.1,我们看到,在分析指数式 $b^t$ 时,$b$ 值可定义为七个不同的区域。但因为模型设定$\sigma$ ,$\beta$,$\delta$ $>0$,这实际上已排除了前两个区域,所以只留下了列在表 16.2 中的五种可能的情况。对于每个区域,第二列对 $b$ 值的设定可以等价地转化为第三列所示的对$\sigma$ 值的设定。例如,在区域 III,$b$ 设定为 $0 < b < 1$,因此,可以写出

$$0 < 1 - \sigma(\beta + \delta) < 1$$

$$-1 < -\sigma(\beta + \delta) < 0 \qquad [三项均减去 1]$$

和 $$\frac{1}{\beta + \delta} > \sigma > 0 \qquad [以 -(\beta + \delta) 通除各项]$$

最后一式给出了所求的对于区域 III 的关于$\sigma$ 值的规定。对于其它区域的变换也可按类似步骤完成。因为在每一区域中,时间路径 的类型早已从图16.1知道,所以,$\sigma$ 值的设定可以使我们由已

知的$\sigma$，$\beta$和$\delta$值识别时间路径$P_t$的一般性质，如表16.2的最后一行所示。

表16.2　时间路径的类型

| 区域 | $b \equiv 1-\sigma(\beta+\delta)$的值 | $\sigma$值 | 时间路径$P_t$的性质 |
|---|---|---|---|
| III | $0<b<1$ | $0<\sigma<\frac{1}{\beta+\delta}$ | 非振荡与收敛 |
| IV | $b=0$ | $\sigma=\frac{1}{\beta+\delta}$ | 保持在均衡状态* |
| V | $-1<b<0$ | $\frac{1}{\beta+\delta}<\sigma<\frac{2}{\beta+\delta}$ | 具有衰减振荡 |
| VI | $b=-1$ | $\sigma=\frac{2}{\beta+\delta}$ | 具有均匀振荡 |
| VII | $b<-1$ | $\sigma>\frac{2}{\beta+\delta}$ | 具有发散振荡 |

568

* 价格在此情况下将保持在均衡状态这一事实，也可以直接由(16.14)看出来。$\sigma=1/(\beta+\delta)$时，$P_t$的系数变为零，(16.14)简化为$P_{t+1}=\sigma(\alpha+\gamma)=(\alpha+\gamma)/(\beta+\delta)=\bar{P}$。

**例1**　若我们模型中的卖者总是按照存货减少(增加)数量的10%来提(降)价，且若需求曲线斜率为$-1$，供给曲线斜率为15(斜率均是相对于价格轴的)，那么，我们将会求出何种时间路径?

这里，我们有$\sigma=0.1$，$\beta=1$，$\delta=15$。因为$1/(\beta+\delta)=1/16$，$2/(\beta+\delta)=1/8$，$\sigma(=1/10)$的值位于上面两个值之间，因此，它属于区域V中的情况。时间路径具有衰减振荡的特征。

## 结果的图示总结

表16.2包含五种关于$\sigma$值设定的可能的情况，如果以图形表示这些结果，则更易于把握。由于$\sigma$值的设定实质上是参数$\sigma$和$(\beta+\delta)$相对大小的比较，所以，我们像图16.3那样，对$(\beta+\delta)$绘出$\sigma$

的图形。注意，我们只关心正的象限，因为根据模型的设定，$\sigma$ 和 $(\beta+\delta)$ 均为正值。由表 16.2 显然可以知道，区域 IV 和 VI 是分别由方程 $\sigma=1/(\beta+\delta)$ 和 $\sigma=2/(\beta+\delta)$ 设定的。因为每个方程都可以绘成一个等轴双曲线，所以这两个区域在图 16.3 中可由两条双曲线来表示。而且一旦我们有了这两条双曲线，其它三个区域便可以马上确定了。比如，区域 III 只是位于下面的那条双曲线的点的集合，在此区域中，$\sigma<1/(\beta+\delta)$。类似地，区域 V 表示位

569

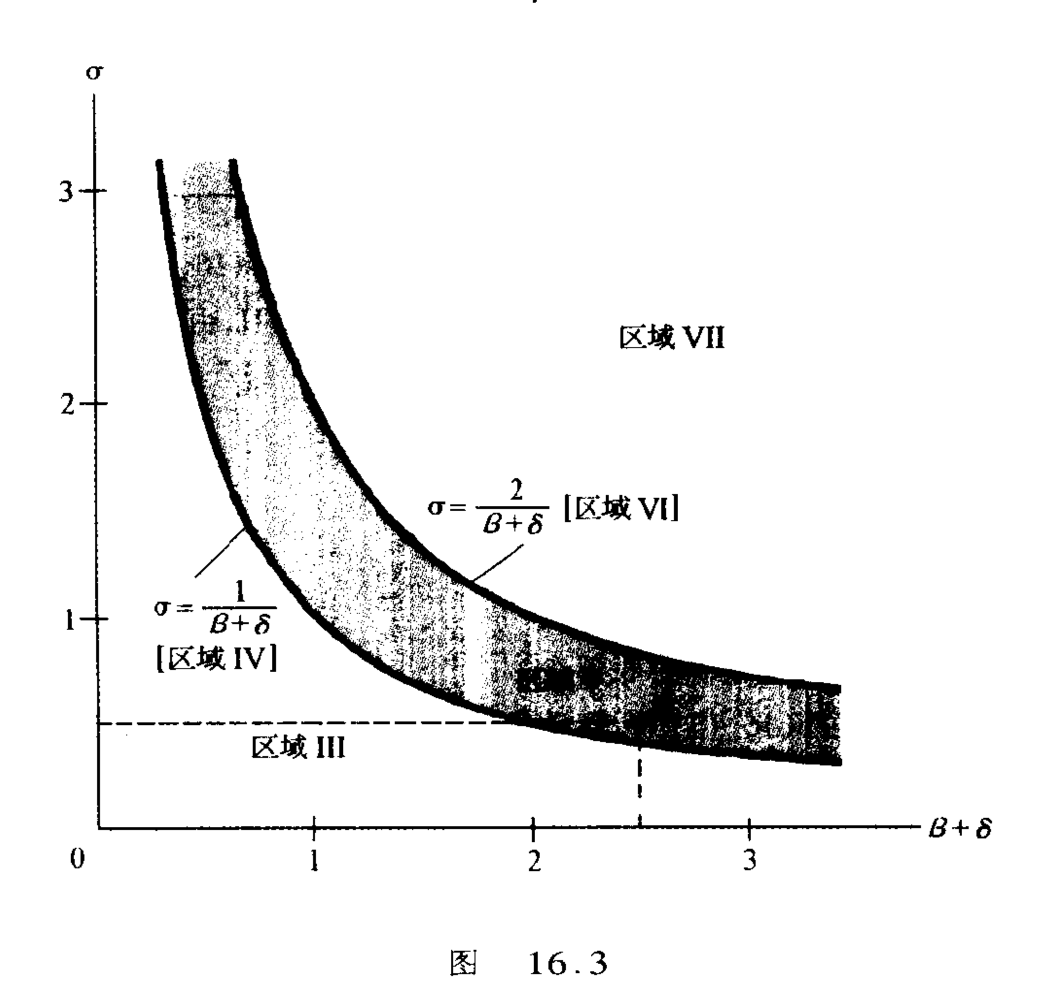



图 16.3

于两条双曲线间的点的集合，而位于上面那条双曲线以上的点则属于区域 VII。

**例 2** 若 $\sigma=1/2$，$\beta=1$，$\delta=3/2$，模型(16.3)含产生收敛的时间路径 $P_t$ 吗？给定的参数值对应于图 16.3 中的点 $A$。因为它位

于区域 V 中，所以路径是振荡收敛的。

读者应注意，在刚才介绍的两个模型中，我们的分析结果在每一情况下均表述成若干可供选择的可能情况——蛛网模型有三种振荡路径，存货模型中则有五种时间路径。当然，这种分析结果的丰富性源于模型的参数形成。而且，我们的分析结果不能用单一的明确的答案来描述是一个优点，而非缺陷。

### 练习 16.5

---

1 在解(16.14)时，为什么运用公式(16.8′)而不用(16.9′)?

2 在表 16.2 的基础上，检验从区域 IV 至区域 VII，由 $b$ 值设定向 $\sigma$ 值设定转换的正确性。

3 若模型(16.13)具有如下数值形式：

$Q_{dt}=21-2P_t$

$Q_{st}=-3+6P_t$

$P_{t+1}=P_t-0.3(Q_{st}-Q_{dt})$

求时间路径 $P_t$，并确定它是否收敛。

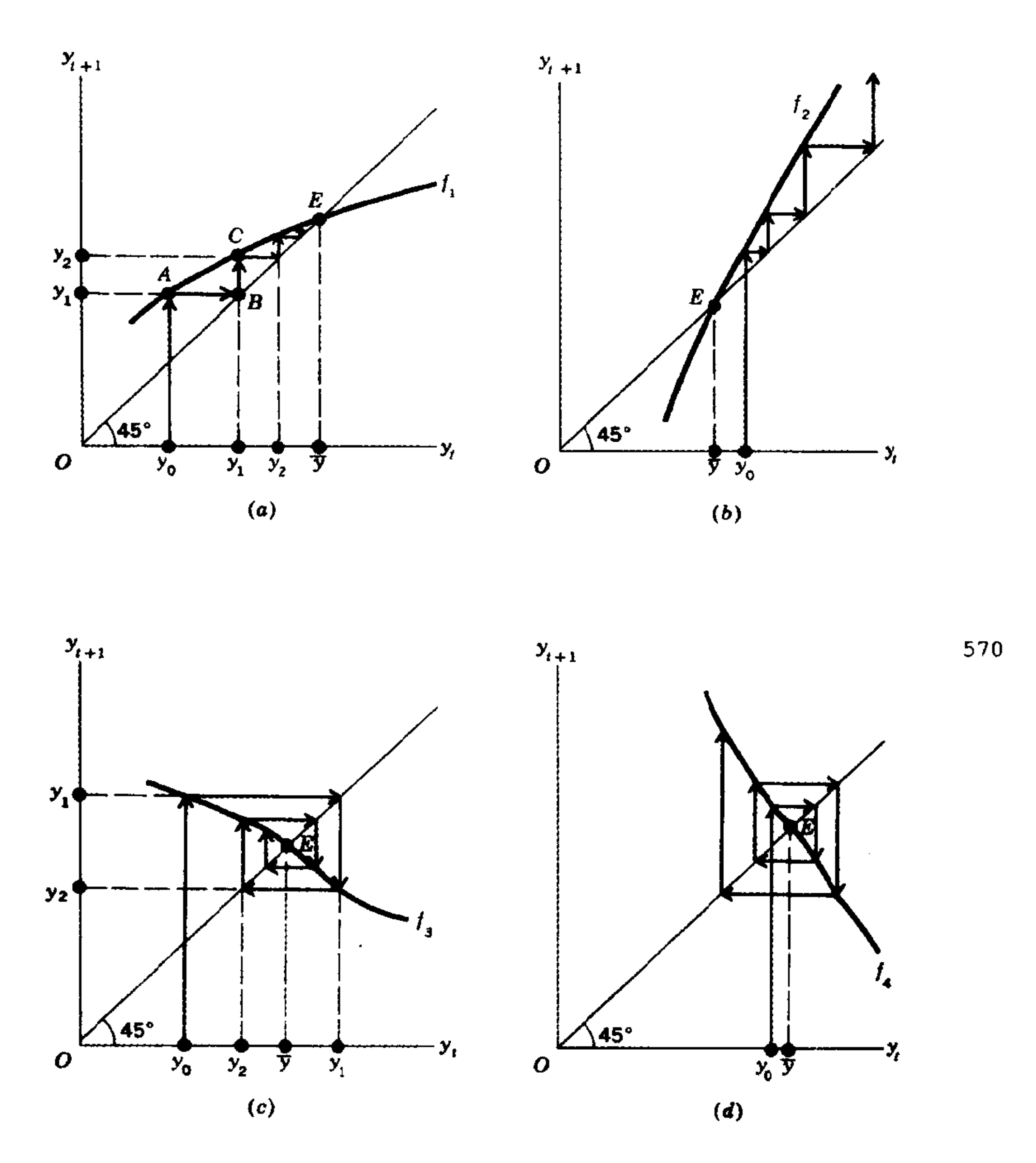
4 假定在模型(16.13)中，假设每期供给是一个固定数量，比如 $Q_{st}=k$，而非价格的函数。分析价格随时间变化的行为。要使解具有经济意义，对 $k$ 值应施加何种限制?

---

## 16.6 非线性差分方程——定性图解法

迄今为止，我们在模型中仅采用线性差分方程的形式。然而，

现实经济生活并不总是符合这种方便的线性形式。幸运的是，当一阶差分方程中出现非线性情况时，仍存在一种在非常一般的情




570

图 16.4

况下就可以应用的容易的分析方法。这种具有图示性质的分析方法与14.6节介绍的一阶微分方程的定性分析方法，有着极大的相似之处。 571

## 相位图

仅有变量 $y_{t+1}$和 $y_t$ 的非线性差分方程,如

$$y_{t+1} + y_t^3 = 5 \qquad 或 \qquad y_{t+1} + \sin y_t - \ln y_t = 3$$

等,可统一以方程

$$y_{t+1} = f(y_t) \tag{16.16}$$

来表示。其中 $f$ 只要仅是 $y_t$ 的函数,复杂到何种程度都无所谓。把两个变量 $y_{t+1}$和 $y_t$ 作为两轴在笛卡尔坐标系中绘出,所产生的图形构成了相位图,对应于 $f$ 的曲线为相位线。由此图便可以通过迭代过程分析变量的时间路径。

这里使用的相位图和相位线这两个术语与微分方程中的情况类似,但在构图时有一点不相似之处要加以注意。在微分方程情况下,我们像在图 14.3 中那样,将 $dy/dt$ 对 $y$ 绘出,因而,为使现在情况与之完全类似,我们应将 $\Delta y_t$ 标在纵轴上,将 $y_t$ 标在横轴上。这样做并非不可能,但如图 16.4 那样,将 $y_{t+1}$置于纵轴上要方便得多。注意,在图 16.4 中的每个图中都绘出了一条 45°线,我们会认识到这条线对我们进行图解分析具有极大的帮助。

我们现在介绍包含于图 16.4$a$ 中的方法,在图 16.4$a$ 中我们已绘出了表示特定差分方程$y_{t+1}=f_1(y_t)$的相位线(标为 $f_1$)。若我们已知初始值 $y_0$(绘在横轴上),所以通过迭代,我们可以求出 $y$ 的所有后期值。首先,因为相位线 $f_1$ 按照方程

$$y_1 = f_1(y_0)$$

将初始值 $y_0$ 映射于 $y_1$。我们可由 $y_0$ 直接向上至相位线,与之交于点 $A$,并在纵轴上读出其高度为 $y_1$。其次,我们设法按方程

$$y_2 = f_1(y_1)$$

将 $y_1$ 映射于 $y_2$。为此,我们必须首先将 $y_1$ 绘在横轴上——与 $y_0$ 的第一次映射过程类似。欲将 $y_1$ 由纵轴转绘至横轴,运用 45°线便很容易完成。45°线的斜率为 1,是到横轴和纵轴距离相等的点[如(2,2)及(5,5)等]的轨迹。因此,要从纵轴转绘 $y_1$,只需由 $y_1$ 绘水平线与 45°线交于点 $B$,然后直接向下拐至横轴,便确定了点 $y_1$。重复这一过程,通过相位线上的点 $C$,我们可将 $y_1$ 映射至 $y_2$,然后再通过 45°线转绘 $y_2$,等等。

现在,迭代的性质已清楚了。我们可以观察到,迭代过程只需如下步骤便可实现:顺着箭头由 $y_0$ 到 $A$(在相位线上),到 $B$(在 572 45°线上),再到 $C$(在相位线上)等等——总是在相位线和 45°线间交替——而无需借助于坐标轴。

### 时间路径的类型

当然,刚才描述的图形迭代同样可以应用于图 16.4 中另外三个图形。实际上,图 16.4 中的四个图描述了相位线的四种基本类型,每条相位线代表不同的时间路径、前两条相位线 $f_1$ 和 $f_2$ 具有斜率为正的特征,其中一条斜率小于 1,而另一条斜率大于 1:

$$0 < f_1'(y_t) < 1 \qquad 和 \qquad f_2'(y_t) > 1$$

而剩下的两条相位线斜率为负。具体地,我们有

$$-1 < f_3'(y_t) < 0 \qquad 和 \qquad f_4'(y_t) < -1$$

在图 16.4 中,$y$ 的瞬时均衡值(即 $\overline{y}$)位于相位线与 45°线的交点上,我们将其标为 $E$。瞬时均衡值之所以位于交点上,是因为相位线上的点 $E$,同时为 45°线上的点,会将 $y_t$ 映射到等值的

$y_{t+1}$；根据定义，当 $y_{t+1}=y_t$ 时，$y$ 必处于瞬时均衡状态。我们的主要任务是确定，给定初始值 $y_0 \neq \bar{y}$，由相位线所隐含的变化模式是会将我们至始至终引向 $\bar{y}$（收敛），还是会引向背离 $\bar{y}$ 的方向（发散）。

对于相位线 $f_1$，迭代过程在一个稳定的路径中由 $y_0$ 引向 $\bar{y}$，没有振荡。读者可以验证，若将 $y_0$ 置于 $\bar{y}$ 的右边，仍会稳定地移向 $\bar{y}$，但方向是向左边移动。这些时间路径收敛于均衡，且其一般图形与图 16.1 区域 III 所示的图形的类型应相同。

相位线 $f_2$ 的斜率大于 1，会出现离散的时间路径。由大于 $\bar{y}$ 的初始值 $y_0$ 出发，箭头总是指向背离均衡的越来越高的 $y$ 值。读者可以验证，小于 $\bar{y}$ 的初始值也会产生类似的稳定的发散运动，但方向同上面相反。

当相位线像 $f_3$ 和 $f_4$ 那样，斜率为负时，振荡将会代替稳定运动，而且现在会出现相对于均衡的过度调整（*over shooting*）现象。在图 $c$ 中，$y_0$ 指向 $y_1$，$y_1$ 大于 $\bar{y}$；紧接着 $y_1$ 的是 $y_2$，而 $y_2$ 又小于 $\bar{y}$，如此等等。在这种情况下，时间路径的收敛性由相位线斜率的绝对值小于 1 所决定。此即相位线 $f_3$ 的情况，这里过度调整的范围在依次各期中逐渐递减。而相位线 $f_4$ 的绝对值大于 1，出现相反的倾向，导致离散的时间路径。573

由相位线 $f_3$ 和 $f_4$ 所生成的振荡时间路径使人联想起图16.2 中的蛛网。但在图 16.4$c$ 或 $d$ 中，蛛网是围绕着相位线（它包含时滞）和 45°线结成的，而不是围绕着需求曲线和（滞后）供给曲线结成的。在这里，45°线用作转绘 $y$ 值的辅助机制，而在图 16.2 中，$D$ 曲线（起类似于图 16.4 中 45°线的作用）是模型本身的整体部

分。具体地说,一旦 $Q_{st}$ 在供给曲线上确定了,我们便让箭头交于 $D$ 曲线,以便找到使"市场出清"时的价格,这是蛛网模型的博弈法则。因而在标出两个轴时存在一个基本不同:在图 16.2 中,存在两个完全不同的变量 $P$ 和 $Q$,但在图 16.4 中,坐标轴表示同一变量 $y$ 在两个紧邻时期的值。但要注意,若我们分析差分方程(16.11)的图形(它构成了蛛网模型),而非(16.10)中单独的需求函数和供给函数,则所得到的图形将是如图 16.4 所示的相位线。换言之,确实存在两种蛛网模型的图形分析方法,且均能得到相等的结果。

由对相位线的上述考察,可得出这样一个基本法则:相位线斜率的代数符号决定是否存在振荡,而其斜率的绝对值则决定收敛性问题。如果某相位线即包括正斜率的部分,也包括负斜率的部分,且如果其斜率的绝对值在某些点大于 1,而在某些点小于 1,那么,时间路径自然会变得非常复杂。但即使在这种情况下,图形迭代分析仍一样容易应用。当然,在迭代过程开始之前,我们必须得到一个初始值。实际上,在那些更为复杂的情况下,不同的初始值可能引致完全不同的时间路径(见练习 16.6-2 和 16.6-3)。

## 具有价格上限的市场

我们现在介绍一个非线性差分方程的例子。在图 16.4 中,四个非线性相位线恰好都是平滑的。在现在的例子中,我们将给出非平滑相位线的例子。

首先,我们取蛛网模型的线性差分方程(16.11),并将其写成

$$P_{t+1}=\frac{\alpha+\gamma}{\beta}-\frac{\delta}{\beta}P_t \quad \left(\frac{\delta}{\beta}>0\right) \tag{16.17}$$

此方程为 $P_{t+1}=f(P_t)$形式，有 $f'(P_t)=-\delta/\beta<0$。在斜率绝对值大于 1，即放大振荡的假设下，我们在图 16.5 中绘出了这个线性相位线。

现在，假设存在一个法定价格上限 $\hat{P}$。它在图 16.5 中表现为一条水平直线，因为无论 $P_t$ 还是 $P_{t+1}$，都不允许超过 $\hat{P}$ 水平。这

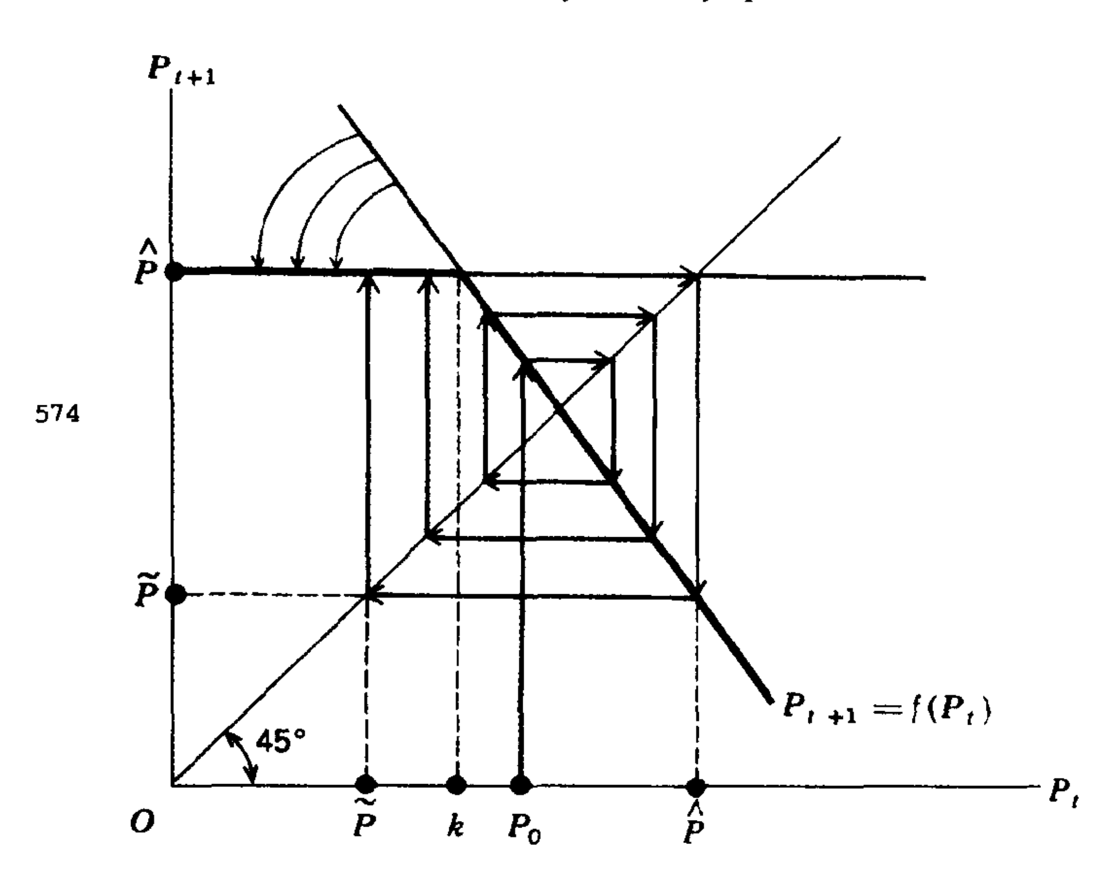



图 16.5

便造成位于 $\hat{P}$ 以上的那部分相位线失效，或者从另一个角度看，这使得 $\hat{P}$ 以上的相位线弯折至 $\hat{P}$ 水平，因而产生了一个“弯折”相位线。[①] 考虑到这个弯折，这条新的(粗)的相位线不仅是非线性

① 严格地讲，我们也应“弯折”水平轴上点 $\hat{P}$ 右边的那部分相位线。但只要另一边已被弯折，像现在这样也没什么不妥，因为 $P_{t+1}$转绘至横轴会自动地将上限 $\hat{P}$ 移至 $P_t$ 轴。

的,而且是非平滑的。像阶梯函数一样,这个弯折线不只需要一个方程来表达?

$$P_{t+1}=\begin{cases}\hat{P} & (\text{对于 } P_t \leqslant k)\\ \dfrac{\alpha+\gamma}{\beta}-\dfrac{\delta}{\beta}P_t & (\text{对于 } P_t > k)\end{cases} \tag{16.17'}$$

其中 $k$ 表示在弯折处的值。

假设初始价格为 $P_0$,我们用图形迭代来求出价格的时间路径。在迭代的第一阶段,当相位线向下倾斜的部分起作用时,必然会表现出放大振荡的倾向。但在几个时期以后,箭头开始接触价格上限,此后,时间路径进入在 $\hat{P}$ 和价格下限 $\tilde{P}$ 之间持久循环运动。因此,由于价格上限,模型固有的放大倾向实际上仍存在,而曾经存在的放大振荡现在则驯化成一种均匀振荡(单位振荡)。

这个结果的意义在于:虽然在线性相位线情况下,当且仅当相位线斜率为 -1 时,才会产生均匀振动,但现在引入非线性之后,即使相位线的斜率不为 -1,仍能产生同样的分析结果。这个结论具有非常重要的经济含义。若有人观测到一个变量的时间路径多 575 少有些均匀振荡,并试图用线性模型来解释它,那么,他不得不依赖于一种非常特殊的——且难以置信的——模型设定,即相位线斜率恰好为 -1。但若引入非线性,无论是平滑还是非平滑的类型,那么便可以采用许多更为合理的假设,每个假设都可给观察到的均匀振荡以合理的解释。

## 练习 16.6

1 在差分方程模型中,变量 $t$ 仅取整数值。这是否意味着在图 16.4 的相位图中,必须将变量 $y_t$ 和 $y_{t+1}$视为离散变量?

2 把一个倒 $U$ 形曲线的左半部分作为一条相位线,并使此倒 $U$ 形曲线与 45°线相交于两点 $L$(左边)和 $R$(右边)。

($a$) 这种情况是多重均衡吗?

($b$) 若初始值 $y_0$ 位于 $L$ 的左边,会得到何种类型的时间路径?

($c$) 若初始值位于 $L$ 和 $R$ 之间呢?

($d$) 若初始值位于 $R$ 的右边呢?

($e$) 对在 $L$ 和 $R$ 的均衡的动态稳定性,分别可以得出什么结论?

3 使用倒 $U$ 形曲线作为相位线。令其向上倾斜的部分与 45°线交于点 $L$,令其向下倾斜的部分与 45°线交于点 $R$。回答上题提出的五个问题。(注意:你的答案取决于绘出相位线的具体方式;要考虑到各种可能性。)

4 在图 16.5 中,废除法定价格上限,增加最低价格 $P_m$。

($a$) 相位线会产生何种变化?

($b$) 它还会是弯折的吗? 还是非线性的吗?

($c$) 还会产生价格的均匀振荡运动吗?

5 参考(16.17′)和图 16.5,证明常数 $k$ 可以表示成

$$k = \frac{\alpha + \gamma}{\delta} - \frac{\beta}{\delta}\hat{P}$$

# 第 17 章　高阶差分方程

576

上一章的经济模型由那些将 $P_t$ 与 $P_{t-1}$ 联系起来的差分方程构成。因为一个时期的 $P$ 值可唯一地确定下一时期的 $P$ 值，所以，一旦初始值 $P_0$ 给定，则 $P$ 的时间路径便完全确定了。但是也可能出现这种情况：$t$ 期的经济变量，比如 $y_t$，不仅取决于 $y_{t-1}$，而且取决于 $y_{t-2}$。这样便引出了二阶差分方程。

严格地讲，二阶差分方程是一个包含表达式 $\Delta^2 y_t$，但不含高于二阶差分的方程。$\Delta^2 y_t$，读作 $y_t$ 的二阶差分。而符号 $\Delta^2$ 是符号 $d^2/dt^2$ 在离散时间情况下的对应物，表示"取二阶差分"如下：

$$\begin{aligned}
\Delta^2 y_t &= \Delta(\Delta y_t) = \Delta(y_{t+1} - y_t) && [\text{由}(16.1)] \\
&= (y_{t+2} - y_{t+1}) - (y_{t+1} - y_t) && [\text{仍由}(16.1)]^{①} \\
&= y_{t+2} - 2y_{t+1} + y_t
\end{aligned}$$

因此，$y_t$ 的二阶差分可以转换为包含两期时滞的项的和。因为像 577 $\Delta^2 y_t$ 和 $\Delta y_t$ 这样的表达式写起来很麻烦，所以我们将二阶差分方程重新定义为包含变量的两期时滞的方程。类似地，三阶差分方程为包含三期时滞的方程；等等。

---

① 即，若我们将 $(y_{t+1} - y_t)$ 中的下标向前移一个时期，得到新表达式 $(y_{t+2} - y_{t+1})$；然后再从后一个式中减去原表达式。注意，因所得的差分可以写成 $\Delta y_{t+1} - \Delta y_t$，所以可以推出如下运算法则：

$$\Delta(y_{t+1} - y_t) = \Delta y_{t+1} - \Delta y_t$$

它使我们联想起可应用于和或差的求导法则。

我们首先集中讨论二阶差分方程的解法，然后再在后面的章节中将其推广至高阶差分方程。为控制讨论的范围，在本章，我们仅讨论常系数线性差分方程。但对常数项和可变项两种形式，均作考察。

## 17.1 具有常系数和常数项的二阶线性差分方程

一类简单的二阶差分方程的形式为

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = c \tag{17.1}$$

读者应注意到，此方程为线性、非齐次，且具有常系数($a_1$, $a_2$)和常数项 $c$ 的差分方程。

### 特别积分

同以前一样，可以预期(17.1)的解具有两部分：表示 $y$ 的瞬时均衡水平的特别积分 $y_p$，表示每一时期与均衡的偏差的余函数 $y_c$。特别积分定义为完备方程的任意解，有时通过试探形式为 $y_t = k$ 的解便可以求出。将 $y$ 的这个常数值(17.1)，我们得到

$$k + a_1 k + a_2 k = c \quad \text{和} \quad k = \frac{c}{1 + a_1 + a_2}$$

因此，只要$(1 + a_1 + a_2) \neq 0$，特别积分为

$$y_p(= k) = \frac{c}{1 + a_1 + a_2} \qquad (a_1 + a_2 \neq -1 \text{ 的情况}) \tag{17.2}$$

**例 1** 求 $y_{t+2} - 3y_{t+1} + 4y_t = 6$ 的特别积分。这里我们有$a_1 =$

$-3, a_2=4$, $c=6$。因为 $a_1+a_2 \neq -1$，由(17.2)可得如下特别积分：

$$y_p = \frac{6}{1-3+4} = 3$$

在 $a_1+a_2=-1$ 情况下，试验解 $y_t=k$ 不成立，所以，必须再试探 $y_t=kt$。将 $y_t=kt$ 代入(17.1)，并记住我们现在有 $y_{t+1}=k(t+1)$, $y_{t+2}=k(t+2)$，我们求得 578

$$k(t+2)+a_1k(t+1)+a_2kt=c$$

和
$$k = \frac{c}{(1+a_1+a_2)t+a_1+2} = \frac{c}{a_1+2}$$

[因为 $a_1+a_2=-1$]

因此，我们可将特别积分写成

$$y_p(=kt) = \frac{c}{a_1+2}t \qquad (a_1+a_2=-1; a_1\neq -2 \text{ 的情况}) \tag{17.2'}$$

**例 2** 求 $y_{t+2}+y_{t+1}-2y_t=12$ 的特别积分。这里，$a_1=1$, $a_2=-2$, $c=12$。显然，不能应用公式(17.2)，但可应用(17.2′)。因此，

$$y_p = \frac{12}{1+2}t = 4t$$

此特别积分表示移动均衡。

若 $a_1+a_2=-1$，但同时 $a_1=-2$（即若 $a_1=-2$, $a_2=1$），则我们可采用形式为 $y_t=kt^2$ 的试探解，这意味着 $y_{t+1}=k(t+1)^2$，等等。如读者可以验证的那样，在此情况下特别积分为

$$y_p = kt^2 = \frac{c}{2}t^2 \qquad (a_1=-2; a_2=1 \text{ 的情况}) \tag{17.2''}$$

但是,因为此公式仅能应用于差分方程为 $y_{t+2}-2y_{t+1}+y_t=c$ 这种情况,所以其价格是非常有限的。

### 余函数

为求出余函数,我们必须集中讨论简化方程

$$y_{t+2}+a_1y_{t+1}+a_2y_t=0 \tag{17.3}$$

解一阶差分方程的经验告诉我们,$Ab^t$ 式在这种方程的通解中起非常重要的作用。因此,我们先试探形式为 $y_t=Ab^t$ 的解,它自然意味着 $y_{t+1}=Ab^{t+1}$,等等。我们的任务便是确定 $A$ 和 $b$ 的值。

将试探解代入(17.3),方程变成

$$Ab^{t+2}+a_1Ab^{t+1}+a_2Ab^t=0$$

或在消去(非零)共同因子 $Ab^t$ 后,有

$$b^2+a_1b+a_2=0 \tag{17.3'}$$

此二次方程——(17.3)或(17.1)的**特征方程**——与(15.4″)具有可比性。它具有两个**特征根**:

$$b_1,b_2=\frac{-a_1\pm\sqrt{a_1^2-4a_2}}{2} \tag{17.4}$$

579 对解 $Ab^t$ 中的 $b$ 而言,上述每个根都是可接受的。事实上,$b_1$ 和 $b_2$ 均应在齐次差分方程(17.3)的通解中出现,恰如在微分方程中的情况一样,此通解必然包括两个**线性无关**的部分,每一部分都有自己的乘积任意常数。

关于特征根,视平方根式(17.4)的情况,可能遇到三种可能的情形。读者会发现,这与15.1节对二阶微分方程的分析极其相似。

**第一种情况(不同的实根)**　当 $a_1^2 > 4a_2$ 时,(17.4)中的平方根为实数,$b_1$ 和 $b_2$ 为不同的实根。在这种情况下,$b_1^t$ 和 $b_2^t$ 线性无关,余函数可以简单地写成 $b_1^t$ 和 $b_2^t$ 的线性组合,即

$$y_c = A_1 b_1^t + A_2 b_2^t \tag{17.5}$$

读者可将其与(15.7)相比较。

**例 3**　求 $y_{t+2} + y_{t+1} - 2y_t = 12$ 的解。此方程的系数为 $a_1 = 1, a_2 = -2$。由(17.4),可求得特征根为 $b_1 = 1, b_2 = -2$。因此,余函数为

$$y_c = A_1(1)^t + A_2(-2)^t = A_1 + A_2(-2)^t$$

因为在例 2 中,给定差分方程的特别积分已求得为 $y_p = 4t$,所以,我们可将通解写成

$$y_t = y_c + y_p = A_1 + A_2(-2)^t + 4t$$

仍有两个任意常数 $A_1$ 和 $A_2$ 有待于确定。要确定这两个值,必须有两个初始条件。假设我们有 $y_0 = 4$ 和 $y_1 = 5$。则依次令通解中的 $t = 0$ 和 $t = 1$,我们求得

$y_0 = A_1 + A_2$ (= 4 由第一个初始条件)

$y_1 = A_1 - 2A_2 + 4$ (= 5 由第二个初始条件)

任意常数可以确定为 $A_1 = 3$, $A_2 = 1$。则最终可将定解写成

$$y_t = 3 + (-2)^t + 4t$$

**第二种情况(重实根)**　当 $a_1^2 = 4a_2$ 时,平方根(17.4)为零,特征根为重根:

$$b(= b_1 = b_2) = -\frac{a_1}{2}$$

现在,若我们将余函数表示为(17.5)的形式,则两部分将合并为一 580

项：

$$A_1 b_1^t + A_2 b_2^t = (A_1 + A_2) b^t \equiv A_3 b^t$$

此式无效，因为现在缺一个常数。

为补上缺失的部分（我们回顾一下，这部分应与 $A_3 b^t$ 项线性无关），还需重施以变量 $t$ 乘 $b^t$ 这个故技。这样这个新的项可取 $A_4 t b^t$ 这种形式。它与 $A_3 b^t$ 线性无关应是很明显的，因为我们永远不能给 $A_3 b^t$ 加上一个常系数而得到 $A_4 t b^t$。$A_4 t b^t$ 像 $A_3 b^t$ 一样，确实可以作为齐次方程(17.3)的合格解这一事实，可以很容易得到验证：只需将 $y_t = A_4 t b^t$〔和 $y_{t+1} = A_4(t+1) b^{t+1}$ 等〕代入(17.3)①，便可以看到(17.3)简化为一个恒等式 0=0。

因此，重根情况下的余函数为

$$y_c = A_3 b^t + A_4 t b^t \tag{17.6}$$

读者可将其与(15.9)相比较。

**例 4** 求 $y_{t+2} + 6y_{t+1} + 9y_t = 4$ 的余函数。系数为 $a_1 = 6, a_2 = 9$，可求得特征根为 $b_1 = b_2 = -3$。这样我们有

$$y_c = A_3(-3)^t + A_4 t(-3)^t$$

若我们再进一步，可很容易求得 $y_p = 1/4$，则给定差分方程的通解为

$$y_t = A_3(-3)^t + A_4 t(-3)^t + \frac{1}{4}$$

若给定两个初始条件，便可以确定 $A_3$ 和 $A_4$ 的值。

**第三种情况（复根）** 在最后一种可能，$a_1^2 < 4a_2$ 情况下，特征根为共轭复根。具体地，根的形式为

---

① 代入时应注意，此时有 $a_1^2 = 4a_2$ 及 $b = -a_1/2$。

$$b_1, b_2 = h \pm vi$$

其中

$$h = -\frac{a_1}{2} \text{ 和 } v = \frac{\sqrt{4a_2 - a_1^2}}{2} \tag{17.7}$$

因此,余函数变成

$$y_c = A_1 b_1^t + A_2 b_2^t = A_1 (h + vi)^t + A_2 (h - vi)^t$$

上式表明,解释 $y_c$ 并不容易。但幸运的是,由于在(15.23′)中给出的棣莫弗定理,此余函数很容易化为三角函数,而三角函数我们已知道如何解释。

根据棣莫弗定理,可以写出

$$(h \pm vi)^t = R^t (\cos\theta t \pm i \sin \theta t)$$

581

由(15.10)可知,其中的 $R$ 值(总取正值)为:

$$R = \sqrt{h^2 + v^2} = \sqrt{\frac{a_1^2 + 4a_2 - a_1^2}{4}} = \sqrt{a_2} \tag{17.8}$$

其中 $\theta$ 为$(0,2\pi)$内的角,以弧度度量。它满足条件

$$\cos\theta = \frac{h}{R} = \frac{-a_1}{2\sqrt{a_2}} \quad \text{和} \quad \sin\theta = \frac{v}{R} = \sqrt{1 - \frac{a_1^2}{4a_2}} \tag{17.9}$$

因此,余函数可以变换如下

$$\begin{aligned} y_c &= A_1 R^t (\cos\theta t + i \sin\theta t) + A_2 R^t (\cos\theta t - i \sin\theta t) \\ &= R^t [(A_1 + A_2)\cos\theta t + (A_1 - A_2) i \sin\theta t] \\ &= R^t (A_5 \cos\theta t + A_6 \sin\theta t) \end{aligned} \tag{17.10}$$

其中我们采用了简写符号

$$A_5 \equiv A_1 + A_2 \quad 和 \quad A_6 \equiv (A_1 - A_2)i$$

余函数(17.10)与其在微分方程中的对应物(15.24′)有两点重要区别。首先,表达式 $\cos\theta t$ 和 $\sin\theta t$ 已取代了原来使用的 $\cos vt$ 和 $\sin vt$。其次,乘积因子 $R^t$(以 $R$ 为底的指数)已取代了自然指数式 $e^{ht}$。总之,我们已由复根的笛卡尔坐标系($h$ 和 $v$)转换到极坐标系($R$ 和 $\theta$)。一旦 $h$ 和 $v$ 已知,则 $R$ 和 $\theta$ 的值可由(17.8)和(17.9)确定。如果我们能够首先确定 $a_1^2 < 4a_2$,且根确实为复根,则也有可能通过(17.8)和(17.9),直接由参数 $a_1$ 和 $a_2$ 计算 $R$ 和 $\theta$ 值。

**例5** 求 $y_{t+2} + 1/4 y_t = 5$ 的通解。这里,系数 $a_1 = 0$ 和 $a_2 = 1/4$,这是一个 $a_1^2 < 4a_2$ 的复根的例子。由(17.7)可知,根的实和虚的部分为 $h = 0$, $v = 1/2$。由(17.8)可得

$$R = \sqrt{0 + (\frac{1}{2})^2} = \frac{1}{2}$$

因为 $\theta$ 值可满足两个方程

$$\cos\theta = \frac{h}{R} = 0 \text{ 和 } \sin\theta = \frac{v}{R} = 1$$

582 由表 15.1 可得出结论

$$\theta = \frac{\pi}{2}$$

因而,余函数为

$$y_c = \left(\frac{1}{2}\right)^t \left(A_5 \cos\frac{\pi}{2}t + A_6 \sin\frac{\pi}{2}t\right)$$

为求 $y_p$,我们在完全方程中尝试常数解 $y_t = k$。这产生 $k = 4$,因此,$y_p = 4$,且通解可以写成

$$y_t = \left(\frac{1}{2}\right)^t \left(A_5 \cos\frac{\pi}{2}t + A_6 \sin\frac{\pi}{2}t\right) + 4 \qquad (17.11)$$

**例6** 求 $y_{t+2}-4y_{t+1}+16y_t=0$。首先,很容易求得特别积分为 $y_p=0$。这意味着通解 $y_t(=y_c+y_p)$将与 $y_c$ 一致。为求得 $y_c$,我们注意到系数 $a_1=-4$ 和 $a_2=16$,确实能够产生复根。因此,我们可直接将 $a_1$ 和 $a_2$ 的值代入(17.8)和(17.9)以得到

$$R=\sqrt{16}=4$$

$$\cos\theta=\frac{4}{2.4}=\frac{1}{2} \quad 和 \quad \sin\theta=\sqrt{1-\frac{16}{4.16}}=\sqrt{\frac{3}{4}}=\frac{\sqrt{3}}{2}$$

根据后两个方程,由表15.2可知

$$\theta=\frac{\pi}{3}$$

由此可知,余函数(这里也是通解)为

$$y_c(=y_t)=4^t\left(A_5\cos\frac{\pi}{3}t+A_6\sin\frac{\pi}{3}t\right) \qquad (17.12)$$

## 时间路径的收敛性

同在一阶差分方程中的情况一样,时间路径 $y_t$ 的收敛性仅取决于当 $t\to\infty$ 时,$y_c$ 是否趋近于零。因此,我们在图16.1中所了解的关于 $b^t$ 式的各种图形仍可应用,尽管在这里我们必须考察**两个**特征根,而非一个特征根。

首先考察不同实根的情况:$b_1\neq b_2$。若 $|b_1|>1$,$|b_2|>1$,则余函数(17.5)中的两项 $A_1b_1^t$ 和 $A_2b_2^t$ 将是放大的,因此 $y_c$ 必然是发散的。相反,若 $|b_1|<1$,$|b_2|<1$,当 $t$ 无限增大时,$y_c$ 中的两项将收敛于零,$y_c$ 也将收敛于零。但若 $|b_1|>1$ 而 $|b_2|<1$,会如何呢?在这种中间情况下,很明显,$A_2b_2^t$ 项将会"消失",而另一项会越来越偏离零值。由此可知,$A_1b_1^t$ 最终必将控制局势,并使

路径发散。

583 我们将绝对值较大的那个根称作强根(*dominant root*)。由此看来,实际决定时间路径的特征,至少是关于其敛散性这一特征的是强根。实际情况也的确如此。因此,我们可以这样表述:无论初始条件如何,当且仅当强根的绝对值小于1时,时间路径将是发散的。读者可以验证,在两个根的绝对值都大于1或小于1的情况下(上面讨论过),以及在一个根的绝对值恰好为1的情况下(上面未曾讨论),这个结论都是成立的。但要注意,尽管收敛性最终仅取决于强根,但非强根也会对时间路径施加一定的影响,至少在起始阶段是如此。因此,$y_t$ 的确切图形仍取决于两个根。

其次考察重根的情况,我们发现如(17.6)所示,余函数包含项 $A_3b^t$ 和 $A_4tb^t$。前者我们早已熟悉,但对后者(它包含一个乘积因子 $t$)仍需做一点解释。如果 $|b|>1$,$b^t$ 项将放大,而乘积项 $t$ 随着 $t$ 的增加,会进一步增强放大性。另一方面,如果 $|b|<1$,则 $b^t$ 部分(当 $t$ 增加时,它趋于零)和 $t$ 部分变化方向相反,即 $t$ 值将会抵销而非强化 $b^t$。那么,哪种力量更强一些呢?答案是,$b^t$ 的衰减力量总是会超过 $t$ 的放大力量。因此,在重根情况下对收敛性的基本要求仍是根的绝对值小于1。

**例7** 分析上面例3和例4的解的收敛性。例3的解为

$$y_t = 3 + (-2)^t + 4t$$

其根分别为1和-2,$[3(1)^t = 3]$,而且这里存在一个移动均衡 $4t$。强根为-2,时间路径是发散的。

对于例4,其解为

$$y_t = A_3(-3)^t + A_4t(-3)^t + \frac{1}{4}$$

其中$|b|=3$，时间路径也是发散的。

现在我们考察复根的情况。由余函数的一般形式(17.10)，

$$y_c = R^t(A_5\cos\theta t + A_6\sin\theta t)$$

显然可知，括号中的表达式，像(15.24′)中的表达式一样，将产生一种周期性波动形式。但因在这里，变量 $t$ 仅取整数值 0,1,2,…，我们仅能捕捉并利用三角函数图形中点的子集。在每个这样的点上，直到达到下一个相关的点以前，$y$ 值在一个完整的时期内 584 都是有效的。如图 17.1 所描述的那样，所产生的路径既不是通常的振荡形式(在紧邻的时期中，不在 $y_p$ 值的上下交替)，也不是通常的波动形式(非平滑)，而是表现出一种阶梯波动。就收敛性而言，尽管决定性的因素实际上是 $R^t$ 项，它像(15.24′)中的 $e^{ht}$ 项一样，将确定阶梯波动在 $t$ 增加时是得到强化，还是受到削弱。在现在这种情况下，当且仅当 $R<1$ 时，波动才能逐渐缩减。因为根据定义，$R$ 是共轭复根$(h \pm vi)$的绝对值，所以，收敛性的条件仍是特征根的绝对值小于 1。

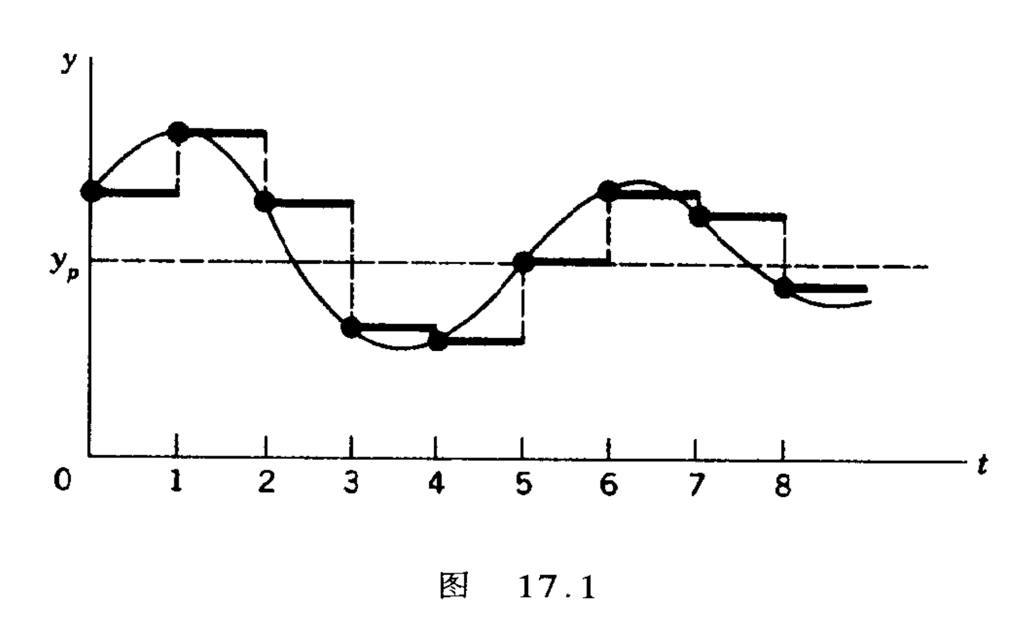



图 17.1

概言之，对于特征根的所有三种情况，无论初始条件为如何，

当且仅当每个根的绝对值小于 1 时,时间路径将会收敛于(一个稳定的或移动的)瞬时均衡。

**例 8** 时间路径(17.11)和(17.12)收敛吗? 在(17.11)中,我们有 $R=1/2$,所以时间路径将收敛于一个稳定均衡(=4)。而在(17.2)中,我们有 $R=4$,所以时间路径不再收敛于均衡(=0)。

### 练习 17.1

---

1 写出下列每个方程的特征方程,并求出特征根:

(a) $y_{t+2}-y_{t+1}+\frac{1}{2}y_t=2$ (c) $y_{t+2}+\frac{1}{2}y_{t+1}-\frac{1}{2}y_t=5$

(b) $y_{t+2}-4y_{t+1}+4y_t=7$ (d) $y_{t+2}-2y_{t+1}+3y_t=4$

2 对于上题中的每个差分方程,根据特征根判定时间路径是否包含振荡或阶梯波动,以及时间路径是否是放大的。

585 3 求练习 17.1-1 中的方程的特别积分。它们表示平稳均衡或移动均衡吗?

4 解下列差分方程:

(a) $y_{t+2}+3y_{t+1}-\frac{7}{4}y_t=9$ $(y_0=6;y_1=3)$

(b) $y_{t+2}-2y_{t+1}+2y_t=1$ $(y_0=3;y_1=4)$

(c) $y_{t+2}-y_{t+1}+\frac{1}{4}y_t=2$ $(y_0=4;y_1=7)$

5 分析上题所得到的时间路径。

---

## 17.2 萨缪尔森乘数—加速相互作用模型

我们引用萨缪尔森教授的经典的*相互作用*模型,作为描述二

阶差分方程在经济学中应用的一个例子。此模型探索当加速原理与凯恩斯乘数一起发生作用时，收入决定的动态过程。[1] 此外，此模型还证明，仅仅是乘数和加速数的相互作用，就能够产生内生的周期性波动。

## 结　构

假设国民收入 $Y_t$ 由三种支出流组成：消费 $C_t$；投资 $I_t$；政府支出 $G_t$。$C_t$ 被看成上期收入 $Y_{t-1}$ 的函数，而非本期收入的函数。为简单起见，假设 $C_t$ 严格地与 $Y_{t-1}$ 成比例。作为一个“引致”变量，投资是消费者现行支出倾向的函数。当然，正是通过这一引致投资，加速原理才得以进入模型。具体地，我们假设 $I_t$ 与消费增量 $\Delta C_{t-1}=C_t-C_{t-1}$ 成固定比例。而第三个支出流 $G_t$，则可视为外生变量。事实上，我们将假设它是一个常数，并以 $G_0$ 表示之。

这些假定可以转换成如下方程组：

$$
\begin{aligned}
Y_t &= C_t + I_t + G_0 \\
C_t &= \gamma Y_{t-1} \qquad (0<\gamma<1) \qquad\qquad (17.13)\\
I_t &= \alpha(C_t - C_{t-1}) \qquad (\alpha>0)
\end{aligned}
$$

其中 $\gamma$（希腊字母 gamma）表示边际消费倾向，$\alpha$ 表示加速数（加速系数的简写）。注意，如果从模型中剔除引致投资，我们就得到一 586 个描述动态乘数过程的一阶差分方程（参见 16.2 节例 2）。但若模型中包含引致投资，我们便得到一个描述乘数与加速数相互作

[1] 保罗·萨缪尔森：“乘数分析与加速原理的相互作用”，载《美国经济评论》，1939 年 5 月，75—78 页。重印于美国经济协会《经济周期理论文摘》，里查德·艾尔文主编，Homewood 出版公司 1944 年版，261—269 页。

用的二阶差分方程。

利用第二个方程，我们可用收入将 $I_t$ 表示如下：

$$I_t = \alpha(\gamma Y_{t-1} - \gamma Y_{t-2}) = \alpha\gamma(Y_{t-1} - Y_{t-2})$$

将此式与 $C_t$ 代入(17.13)中的第一个方程并整理，模型可以化简为一个方程

$$Y_t - \gamma(1+\alpha)Y_{t-1} + \alpha\gamma Y_{t-2} = G_0$$

或者等价地(将下标前移两个时期)，

$$Y_{t+2} - \gamma(1+\alpha)Y_{t+1} + \alpha\gamma Y_t = G_0 \qquad (17.14)$$

由于它是一个具有常系数和常数项的二阶线性差分方程，所以可用刚才学过的方法解之。

## 解　法

作为特别积分，由(17.2)，我们有

$$Y_p = \frac{G_0}{1-\gamma(1+\alpha)+\alpha\gamma} = \frac{G_0}{1-\gamma}$$

读者也许注意到，表达式 $1/(1-\gamma)$ 只是一个乘数，但它只是在不存在引致投资时才成立。因此，$G_0/(1-\gamma)$(外生支出乘以乘数)应在下述意义上给出均衡收入：此收入水平满足均衡条件"国民收入=总支出"〔参见(3.24)〕。然而，作为此模型的特别积分，它也给出瞬时均衡收入。

关于余函数，存在三种可能的情况。在这里，第一种情况($a_1^2 > 4a_2$)的特征为

$$\gamma^2(1+\alpha)^2 > 4\alpha\gamma \quad \text{或} \quad \gamma(1+\alpha)^2 > 4\alpha$$

或

$$\gamma > \frac{4\alpha}{(1+\alpha)^2}$$

类似地，要描述第二、三种情况的特征，我们只需将上面最后一个不等式中的>号分别变成=号和<号即可。在图 17.2 中，我们绘出了方程$\gamma = 4\alpha/(1+\alpha)^2$ 的图形。根据上面的讨论，恰好位于此曲线上的($\alpha,\gamma$)数偶属于第二种情况。而位于该曲线**上面**(包含较大的$\gamma$ 值)的($\alpha,\gamma$)数偶属于第一种情况，位于该曲线**下面**的($\alpha,\gamma$)数偶属于第三种情况。

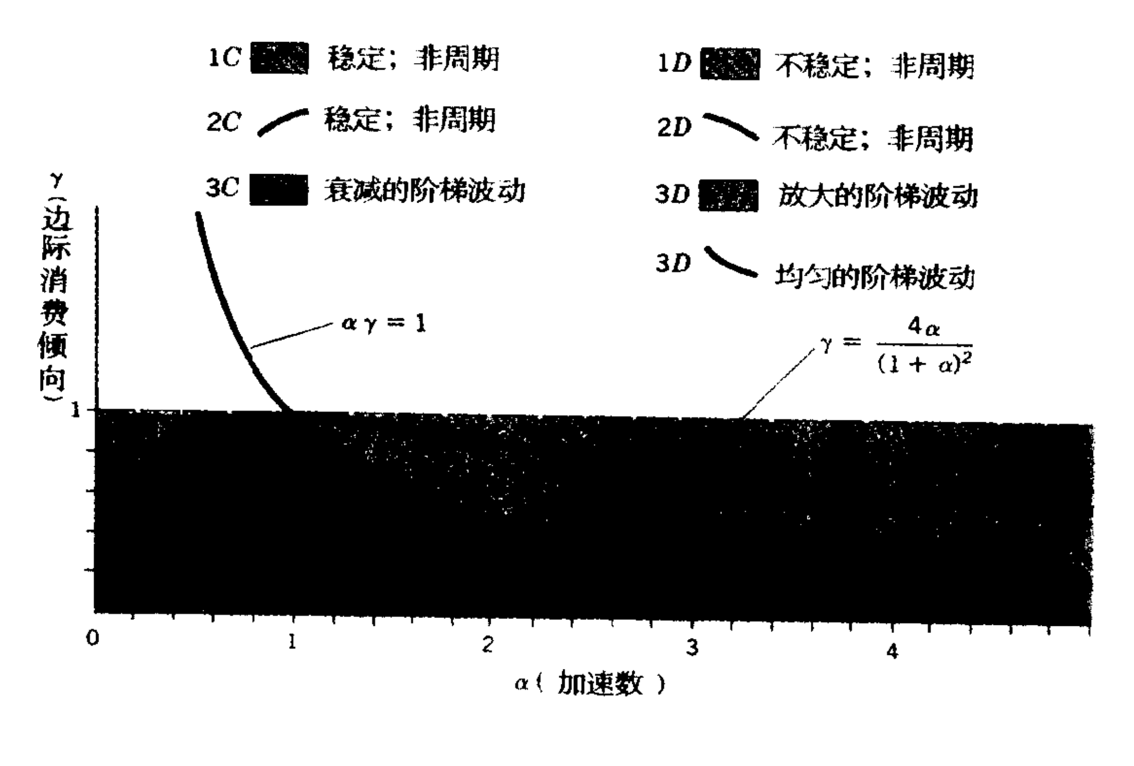



图 17.2

这种具有图 17.2 的图形表示的三重分类是重要的，因为它清楚地揭示了这样一些条件，在此条件下乘数与加速数的相互作用可内生地产生周期性波动。但这种分类并未谈及 $Y$ 的时间路径的敛散性。因此，在每一情况下，我们还需要区分**衰减**与**放大**两种子情况。当然，我们可以通过引用一些数字例子来简单地说明这种子情况，这是处理这一问题的简单方式。不过我们还是设法求

587

出收敛性和发散性的一般条件;尽管这很麻烦,但却更有价值。

## 收敛性与发散性

差分方程(17.14)具有特征方程

$$b^2 - \gamma(1+\alpha)b + \alpha\gamma = 0$$

它产生两个根

$$b_1, b_2 = \frac{\gamma(1+\alpha) \pm \sqrt{\gamma^2(1+\alpha)^2 - 4\alpha\gamma}}{2}$$

因为收敛性与发散性取决于 $b_1$ 和 $b_2$ 的值,又因为 $b_1$ 和 $b_2$ 值取决于参数 $\alpha$ 和$\gamma$ 的值,所以,收敛与发散的条件应当可以用 $\alpha$ 和$\gamma$ 588 值表示。为此,我们可以利用这一事实:由(15.6),两个特征根总可以通过如下两个方程联系起来:

$$b_1 + b_2 = \gamma(1+\alpha) \tag{17.15}$$

$$b_1 b_2 = \alpha\gamma \tag{17.15'}$$

在这两个方程的基础上,我们可以观察到

$$\begin{aligned}(1-b_1)(1-b_2) &= 1-(b_1+b_2)+b_1b_2 \\ &= 1-\gamma(1+\alpha)+\alpha\gamma = 1-\gamma\end{aligned} \tag{17.16}$$

鉴于模型设定 $0<\gamma<1$,有必要对这两个根施加条件

$$0 < (1-b_1)(1-b_2) < 1 \tag{17.17}$$

现在,我们来考察第一种情况下的收敛性问题,其中两个根为不同的实根。因为根据假设,$\alpha$ 和$\gamma$ 均为正,(17.15′)表明 $b_1 b_2 >0$,这意味着 $b_1$ 和 $b_2$ 具有相同的代数符号。进而,因为$\gamma(1+\alpha)>0$,所以,(17.15)表明 $b_1$ 和 $b_2$ 必为正。因此,在第一种情况下,时间路径 $Y_t$ 不会产生振荡。

尽管已知 $b_1$ 和 $b_2$ 的符号，但在第一种情况下至少存在 5 种 $(b_1, b_2)$ 值的组合，每种组合关于 $\alpha$ 和$\gamma$ 的对应值如下：

$$(i)\ 0<b_2<b_1<1 \quad \Rightarrow \quad 0<\gamma<1;\alpha\gamma<1$$

$$(ii)\ 0<b_2<b_1=1 \quad \Rightarrow \quad \gamma=1$$

$$(iii)\ 0<b_2<1<b_1 \quad \Rightarrow \quad \gamma>1$$

$$(iv)\ 1=b_2<b_1 \quad \Rightarrow \quad \gamma=1$$

$$(v)\ 1<b_2<b_1 \quad \Rightarrow \quad 0<\gamma<1;\alpha\gamma>1$$

$i$ 可能性（其中 $b_1$ 和 $b_2$ 为正分数）完全满足条件(17.17)，并与模型设定 $0<\gamma<1$ 一致。在此可能性下，两根之积必然也为正分数，由(17.15′)，这意味着 $\alpha\gamma<1$。相反，下面的三种可能性都违背条件(17.17)，并产生不可接受的$\gamma$ 值（参见练习 17.2－3），因此，必须将它们排除掉。但可能性 $v$ 是可接受的。由于 $b_1$ 和 $b_2$ 均大于 1，(17.17)仍然得到满足，但这次，由(17.15′)，我们有 $\alpha\gamma>1$（而非 $\alpha\gamma<1$）。结果在第一种情况下，只有两种可接受的子可能性。第一种子可能性（可能性 $i$）包含分数根 $b_1$ 和 $b_2$，因而产生了 $Y$ 的一个收敛时间路径。另一种子情况（可能性 $v$）的根大于 1，因而产生一个发散的时间路径。但就 $\alpha$ 和$\gamma$ 的值而言，收敛性与发散性的问题仅取决于 $\alpha\gamma<1$ 还是 $\alpha\gamma>1$。这个结论概括 589 在表 17.1 中最上面的部分，其中收敛的子情况标为 1C，发散的子情况标为 1D。

对于第二种情况——重根情况的分析，实质上是类似的。现在根为 $b=\gamma(1+\alpha)/2$，其符号为正，因为 $\alpha$ 和$\gamma$ 均为正。因此仍然不存在振荡。这里我们只需将 $b$ 值分为三种可能性：

$$(vi)\ 0<b<1 \quad \Rightarrow \quad \gamma<1;\alpha\gamma<1$$

$$(vii)\ b=1 \quad \Rightarrow \quad \gamma=1$$

$$(viii)\ b>1 \quad \Rightarrow \quad \gamma<1;\alpha\gamma>1$$

在可能性 $vi$，$b(=b_1=b_2)$为正分数，因此，关于 $\alpha$ 和$\gamma$ 的含义与第一种情况下可能性 $i$ 的情形完全一致。与此类似，可能性 $viii$（其 $b(=b_1=b_2)$大于 1）与可能性 $v$ 的结果相同。而可能性 $vii$ 违背(17.17)，必须被排除。所以只有两种可接受的子情况。第一种子情况（可能性 $vi$）产生一个收敛的时间路径，而另一种子情况（可能性 $viii$）则产生一个发散的时间路径。关于 $\alpha$ 和$\gamma$，收敛与发散的子情况仍然是分别与 $\alpha\gamma<1$ 和 $\alpha\gamma>1$ 相联系的。这些结论列在表 17.1 的中部，其中两种子情况分别标为 2C(收敛)和 2D(发散)。

**表 17.1　萨缪尔森模型的各种可能情形**

| 情况 | 子情况 | $\alpha$ 和$\gamma$ 的值 | 时间路径 $Y_t$ |
|---|---|---|---|
| 1 不同的实根 | | | |
| $\gamma>\frac{4\alpha}{(1+\alpha)^2}$ | $1C:0<b_2<b_1<1$ | $\alpha\gamma<1$ | 非振荡与非波动 |
| | $1D:1<b_2<b_1$ | $\alpha\gamma>1$ | |
| 2 重实根 | | | |
| $\gamma=\frac{4\alpha}{(1+\alpha)^2}$ | $2C:0<b<1$ | $\alpha\gamma<1$ | 非振荡与非波动 |
| | $2D:b>1$ | $\alpha\gamma>1$ | |
| 3 复根 | | | |
| $\gamma<\frac{4\alpha}{(1+\alpha)^2}$ | $3C:R<1$ | $\alpha\gamma<1$ | 具有阶梯波动 |
| | $3D:R\geq 1$ | $\alpha\gamma\geq 1$ | |

最后，在第三种（复根）情况下，我们得到阶梯波动，因而具有内生的商业周期。在此情况下，我们应当考察绝对值 $R=\sqrt{a_2}$ [见(17.8)]作为判定收敛性与发散性的线索，而 $a_2$ 则是差分方程(17.1)中 $y_t$ 项的系数。在本模型中，我们有 $R=\sqrt{\alpha\gamma}$，它产生如下三种可能性：

$$(ix)\ R<1 \Rightarrow \alpha\gamma<1$$
$$(x)\ R=1 \Rightarrow \alpha\gamma=1$$
$$(xi)\ R>1 \Rightarrow \alpha\gamma>1$$

尽管上述几种可能性都是可接受的(参见练习 17.2－4),但仅有 $R<1$ 这种可能性具有收敛的时间路径,在表 17.1 中列为子情况 3C。而另外两种情况在表 17.1 中一并标为子情况 3D。

总之,由表 17.1 我们可以得出结论:当且仅当 $\alpha\gamma<1$ 时,可以得到收敛的时间路径。

## 用图形小结上述分析结果

上述分析采用了较为复杂的情况与子情况的分类方式。如果我们有更为直观的分类图形表示,则会有很大的帮助。图 17.2 便提供了这种图形表示。

在图 17.2 中,模型中所有可接受的有序偶$(\alpha,\gamma)$的集合以各种不同的阴影矩形面积来表示。因为要排除$\gamma=0$ 和$\gamma=1$ 的值,正如要排除到 $\alpha=0$ 的值一样,所以阴影的面积是一种无边的矩形。我们已绘出方程$\gamma=4\alpha/(1+\alpha)^2$ 的图形,以区分表 17.1 中的三种主要情况:在曲线上的点属于第二种情况;位于曲线北面上的点(表示较大的$\gamma$ 值)属于第一种情况;位于曲线南面的点(表示较小的$\gamma$ 值)则属于第三种情况。为区别收敛与发散的子情况,我们现在加上 $\alpha\gamma=1$ 的图形(等轴双曲线),作为另一条分界线。位于该等轴双曲线北面的点满足不等式 $\alpha\gamma>1$,而位于该曲线下面的点则对应于 $\alpha\gamma<1$。这样就可能很容易区分子情况了。在第一种情况下,位于双曲线下面的虚线阴影区域,对应于子情况 1C,而实线阴影区域则与子情况1D相联系。在第二种情况下,即点

位于曲线$\gamma = 4\alpha/(1+\alpha)^2$的情况下，子情况 2C 包括该曲线向上倾斜的部分，而子情况 2D 则对应该曲线向下倾斜的部分。最后，对于第三种情况，等轴双曲线用于区分小点阴影区域(子情况 3C)和小石子阴影区域(子情况 3D)。读者应注意到，后者也包含位于等轴双曲线上的点本身，因为设定的是弱不等式 $\alpha\gamma \geqslant 1$。

因为图 17.2 包含了模型中所有定性的结论，所以如果给定任意有序偶$(\alpha,\gamma)$，通过在图形中绘出该有序偶，我们总可以在图形上找到正确的子情况。

**例 1** 若加速数为 0.8，边际消费倾向为 0.7，会产生何种相互作用的时间路径？有序偶(0.8,0.7)位于小点阴影区域内，属于子情况 3C，因此时间路径以衰减的阶梯波动为特征。

591 **例 2** $\alpha = 2$ 和$\gamma = 0.5$ 表明哪类相互作用的时间路径？有序偶(2,0.5)恰好位于等轴双曲线上，属于子情况 3D。Y 的时间路径仍表现出阶梯波动，但它既非放大，亦非衰减。将其与均匀振荡和均匀波动的情形相比较，我们可以将这种情形称之为“均匀阶梯波动”。但是，后一种情况下的均匀特征一般不能期望它是完美的，因为类似于图 17.1 中的图形，我们只能采纳那些对应于 $t$ 的整数值的在正弦或余弦曲线上的点，但在每一波动周期中，这些 $t$ 值可能会碰到曲线上完全不同的点。

## 练习 17.2

1 参见图 17.2，求下列每组 $\alpha$ 和$\gamma$ 值所属的子情况，并定性地描述相互作用的时间路径：

(a) $\alpha=3.5$; $\gamma=0.8$　　(c) $\alpha=0.2$; $\gamma=0.9$

(b) $\alpha=2$; $\gamma=0.7$　　(d) $\alpha=1.5$; $\gamma=0.6$

2　由上题(a)和(c)部分所给出的 $\alpha$ 和$\gamma$ 值，求每一情况下特征根的数值，并分析时间路径的性质。你的结果与前面获得的结果一致吗？

3　验证第一种情况下的可能性 ii，iii 和 iv，意味着不可接受的$\gamma$ 值。

4　证明在第三种情况下，我们不会遇到$\gamma \geqslant 1$ 的情形。

---

## 17.3　离散时间条件下的通货膨胀与失业

前面在连续时间框架下讨论的通货膨胀与失业的相互作用，也可以在离散时间条件下加以表达。在本节，我们将使用基本相同的假设，描述如何将模型重构为差分方程模型。

### 模　型

前面的连续时间公式(15.5 节)由三个微分方程构成

$$p=\alpha-T-\beta U+h\pi \quad \text{[预期增加的菲力普斯关系]} \tag{15.33}$$

$$\frac{d\pi}{dt}=j(p-\pi) \quad \text{[适应性预期]} \tag{15.34}$$

$$\frac{dU}{dt}=-k(m-p) \quad \text{[货币政策]} \tag{15.35}$$

三个内生变量均为现值：$p$(实际通货膨胀率)，$\pi$(预期通货膨胀率)，$U$(失业率)。在模型中出现 6 个参数，参数 $m$(名义货币增长率，或者货币扩张率)与其它参数的不同之处在于其大小是由政策决定的。592

当把上述方程纳入时期分析模式时，菲利普斯关系(15.33)变

成

$$p_t = \alpha - T - \beta U_t + h\pi_t \quad (\alpha,\beta>0;0<h\leqslant 1) \tag{17.18}$$

在适应性预期方程中，导数必然为差分方程所取代：

$$\pi_{t+1} - \pi_t = j(p_t - \pi_t) \quad (0 < j \leqslant 1) \tag{17.19}$$

同理，货币政策也将变成[①]

$$U_{t+1} - U_t = -k(m - p_{t+1}) \quad (k > 0) \tag{17.20}$$

这三个方程构成了通货膨胀——失业模型的新形式。

### 以 *p* 为变量的差分方程

作为分析新模型的第一步，我们仍设法将模型化简为一个具有单一变量的方程。令该变量为 $p$。相应地，我们把注意力集中于(17.18)。但是，因为(17.18)不同于其它两个方程，它本身不能描述一种变化模式，因而需要我们来创造这样一种模式。我们可以通过对 $p_t$ 取差分，即取 $p_t$ 的一阶差分来做到这一点。根据定义

$$\Delta p_t \equiv p_{t+1} - p_t$$

取一阶差分需要两个步骤：首先，将(17.18)中的时间下标前移一个时期，得到

$$p_{t+1} = \alpha - T - \beta U_{t+1} + h\pi_{t+1} \tag{17.18'}$$

然后从(17.18′)中减去(17.18)，以得到 $p_t$ 的一阶差分，它能够描述所需的变化模式：

---

① 我们曾假定，$U_t$ 的变化取决于$(t+1)$期的实际货币增长率。另一种选择是，有可能使其取决于 $t$ 期的实际货币增长率$(m - p_t)$(参见练习 17.3－4)。

$$p_{t+1}-p_t=-\beta(U_{t+1}-U_t)+h(\pi_{t+1}-\pi_t)$$
$$=\beta k(m-p_{t+1})+hj(p_t-\pi_t) \quad [由(17.20)和(17.19)] \tag{17.21}$$

注意,在(17.21)的第二行,(17.19)和(17.20)给出的另外两个变量的变化模式已被纳入到 $p$ 变量的变化模式中去了。因此,(17.21)现在包容了本模型的所有信息。

但是,$\pi_t$ 项对 $p$ 的研究无关紧要,需将其从上述方程中剔除。593
为此,我们利用这一事实

$$h\pi_t=p_t-(\alpha-T)+\beta U_t \quad [由(17.18)] \tag{17.22}$$

将其代入(17.21),合并同类项,得到

$$(1+\beta k)p_{t+1}-[1-j(1-h)]p_t+j\beta U_t=\beta km+j(\alpha-T) \tag{17.23}$$

但现在又出现了一个有待于删除的 $U_t$ 项。为此,差分(17.23)以得到$(U_{t+1}-U_t)$项,然后再利用(17.20)消去$(U_{t+1}-U_t)$。只有经过这样一个冗长的代换过程,我们才能得到所求的仅含 $p$ 变量的差分方程,经过正规化后,其形式为

$$p_{t+2}-\underbrace{\frac{1+hj+(1-j)(1+\beta k)}{1+\beta k}}_{a_1}p_{t+1}+\underbrace{\frac{1-j(1-h)}{1+\beta k}}_{a_2}p_t=\underbrace{\frac{j\beta km}{1+\beta k}}_{c}$$

## $p$ 的时间路径

由(17.24)的特别积分给出的 $p$ 的瞬时均衡值为

$$\bar{P}=\frac{c}{1+a_1+a_2}=\frac{j\beta km}{\beta kj}=m \quad [由(17.2)]$$

因此,同连续时间模型中的情况一样,均衡的通货膨胀率恰好等于货币扩张率。

至于余函数,依 $a_1^2$ 和 $4a_2$ 的相对大小而定,会产生不同的实根(第一种情况),重根(第二种情况),或者复根(第三种情况)。在本模型中

$$a_1^2 \gtreqless 4a_2 \text{ iff } [1 + hj + (1 - j)(1 + \beta k)]^2$$
$$\gtreqless 4[1 - j(1 - h)](1 + \beta k) \qquad (17.25)$$

比如,如果 $h = 1/2, j = 1/3, \beta k = 5$,则 $a_1^2 = \left(5\frac{1}{6}\right)^2$,而 $4a_2 = 20$,那么便会出现第一种情况。但若 $h = j = 1$,则 $a_1^2 = 4$,而 $4a_2 = 4(1 + \beta k) > 4$,那么便会产生第三种情况。然而,考虑到本模型中包含较多的参数,不可能像萨缪尔森模型中那样,构建一个如 17.2 那样的分类图形。

不过,收敛性的分析仍可按上一节的路线来进行。具体地,由(15.6)可知,两个特征根 $b_1$ 和 $b_2$ 必定满足下列两种关系

$$b_1 + b_2 = -a_1 = \frac{1 + hj}{1 + \beta k} + 1 - j > 0 \qquad [见(17.24)] \quad (17.26)$$

$$b_1 b_2 = a_2 = \frac{1 - j(1 - h)}{1 = + \beta k} \in (0,1) \qquad (17.26')$$

594 进而,在本模型中我们有

$$(1 - b_1)(1 - b_2) = 1 - (b_1 + b_2) + b_1 b_2 = \frac{\beta jk}{1 + \beta k} > 0 \qquad (17.27)$$

现在考察两个根 $b_1$、$b_2$ 为不同实根的情况。因为积 $b_1 b_2$ 为正,所以 $b_1$ 与 $b_2$ 必取相同的符号。进而,因为 $b_1$ 与 $b_2$ 的和为正,所以它们必然均为正,这意味着不会产生振荡。由(17.27)我们可以推断 $b_1$ 与 $b_2$ 均不等于 1,否则 $(1 - b_1)(1 - b_2)$ 将会等于 0,与不等式所表明的含义相违。这表明,按照在萨缪尔森模型中

所列举的关于$(b_1,b_2)$组合的各种可能性,这里不会出现可能性 *ii* 和 *iv*。一个根大于 1,另一个根小于 1 的情形也是不可接受的,否则,$(1-b_1)(1-b_2)$将为负。因此可能性 *iii* 也被排除了。由此可知,$b_1$ 与 $b_2$ 或者二者均大于 1,或者均小于 1。然而,若 $b_1>1$,$b_2>1$(可能性 *v*),将违背(17.26′),结果,最终只有可能性 *i*,即 $b_1$、$b_2$ 均为正分数,从而 $p$ 的时间路径为收敛的情况能够存在。

对第二种情况的分析并没有什么根本的不同。通过同样的推理,我们可以断定重根 $b$ 在本模型中只能为正分数;即可能性 *vi* 是可接受的,但可能性 *vii* 和 *viii* 则不成立。在第二种情况下,$p$ 的时间路径依然是非振荡且收敛的。

至于第三种情况,收敛性要求 $R$(复根的绝对值)小于 1。根据(17.8),$R=\sqrt{a_2}$。由于 $a_2$ 为正分数〔见(17.26′)〕,所以确实有 $R<1$。因此,在第三种情况下,$p$ 的时间路径也是收敛的,尽管这次会出现阶梯波动。

## 关于 $U$ 的分析

如果要分析失业率的时间路径,我们可以以(17.20)为出发点。为排除方程中的 $p$ 项,我们首先代(17.18′)以得到

$$(1+\beta k)U_{t+1}-U_t=k(\alpha-T-m)+kh\pi_{t+1} \tag{17.28}$$

其次,为代换另一个方程作准备,我们对(17.28)取差分以求得

$$(1+\beta k)U_{t+2}-(2+\beta k)U_{t+1}+U_t=kh(\pi_{t+2}-\pi_{t+1}) \tag{17.29}$$

鉴于方程右边存在$\pi$ 的差分表达式,我们可以用一个前移形式的适应性预期方程来代替它。结果

$$(1+\beta k)U_{t+2}-(2+\beta k)U_{t+1}+U_t$$
$$=khj(p_{t+1}-\pi_{t+1}) \qquad (17.30)$$

本模型中的所有信息均包含在此方程中。

595 然而,在适当的关于 $U$ 的差分方程产生以前,我们必须先剔除 $p$ 和$\pi$ 变量。为此,由(17.20)我们注意到

$$kp_{t+1}=U_{t+1}-U_t+km \qquad (17.31)$$

进而,以$(-kj)$通乘(17.22),并前移时间下标,我们可以写成

$$-kjh\pi_{t+1}=-kjp_{t+1}+kj(\alpha-T)-\beta kjU_{t+1}$$
$$=-j(U_{t+1}-U_t+km)+kj(\alpha-T)-\beta kjU_{t+1}$$
[由(17.31)]
$$=-j(1+\beta k)U_{t+1}+jU_t+kj(\alpha-T-m)$$
$$(17.32)$$

这两个结果以 $U$ 变量来表示 $p_{t+1}$ 和$\pi_{t+1}$,可以使我们将其代入(17.30),最终得到仅有 $U$ 变量的差分方程:

$$U_{t+2}-\frac{1+hj+(1-j)(1+\beta k)}{1+\beta k}U_{t+1}+\frac{1-j(1-h)}{1+\beta k}U_t$$
$$=\frac{kj[\alpha-T-(1-h)m]}{1+\beta k} \qquad (17.33)$$

值得注意的是,方程左边的两个常系数与 $p$ 的差分方程〔比如(17.24)〕中的常系数是一致的。因此,前面关于 $p$ 路径余函数的分析同样可以应用到这里。但(17.33)右边的常数项确实有别于(17.24)中的常数项。结果两种情况下的特别积分便不相同。这也应当如此,除了巧合因素外,并无内在的原因可以预期瞬时的均衡失业率与均衡通货膨胀率相同。

## 长期菲利普斯关系

我们已经验证,瞬时均衡失业率为

$$\bar{U}=\frac{1}{\beta}[\alpha-T-(1-h)m]$$

但因已求得均衡通货膨胀率为 $\bar{p}=m$,我们可以通过如下方程将 $\bar{U}$与 $\bar{p}$联系起来

$$\bar{U}=\frac{1}{\beta}[\alpha-T-(1-h)\bar{p}] \tag{17.34}$$

因为此方程仅与均衡的失业率和通货膨胀率有关,所以我们认为它描述了长期的菲利普斯关系。

(17.34)的一个特例引起了经济学家的广泛关注:即 $h=1$ 的情况。若 $h=1$,则 $p$ 项的系数为零,因而会从方程中消失。换言之,$\bar{U}$将变成 $\bar{p}$的常函数。在标准的菲利普斯图形中(失业率被绘成横轴),这个结果产生了一个垂直的菲利普斯曲线。在此情况下 596 的 $\bar{U}$值,被称为自然失业率,它与均衡的通货膨胀率是一致的。这个结论具有明显的政策含义:在长期中,通货膨胀与失业这对孪生魔鬼并不存在像短期中所存在的那种替代关系。

但若 $h<1$ 会如何呢? 在此情况下,(17.34)中 $\bar{p}$ 的系数为负,长期的菲利普斯曲线将向下倾斜,因而通货膨胀与失业之间存在替代关系。因此长期菲利普斯曲线是垂直的,还是斜率为负,关键取决于参数 $h$ 的值。按预期增加的菲利普斯关系,$h$ 度量预期的通货膨胀率与工资结构和实际通货膨胀率相一致的程度。所有这些读者应当是非常熟悉的,因为在 15.5 节例 1 中我们曾讨论过这一问题,而且读者在练习 15.5—4 中也曾处理过这类问题。

## 练习 17.3

---

1 给出由(17.23)推导出(17.24)的中间步骤。

2 证明若本节所讨论的模型化为变量$\pi$ 的差分方程，除了以$\pi$ 替代 $p$ 外，结果与(17.24)将会是相同的。

3 在本节讨论的模型中，$p$ 和 $U$ 的时间路径总是收敛的。如果我们抛掉 $h\leqslant 1$ 的假设，能产生发散的时间路径吗？如果答案是肯定的，那么，在第一、二、三种情况下，哪种发散的“可能性”是可行的？

4 保留方程(17.18)和(17.19)，但将(17.20)变为

$$U_{t+1}-U_t=-k(m-p_t)$$

($a$) 导出变量 $p$ 的新的差分方程。

($b$) 新的差分方程能产生不同的 $\bar{p}$ 吗？

($c$) 假定 $j=h=1$。求特征根分别属于第一、二、三种情况的条件。

($d$) 令 $j=h=1$。当 $\beta k=3,4$ 和 5 时，分别描述 $p$ 的时间路径(包括收敛性或发散性)。

---

# 17.4 推广到可变项和高阶方程

现在，我们准备将我们的方法向两个方向推广：一是向可变项方向推广，二是向高阶差分方程方向推广。

597 **形式为 $cm^t$ 的可变项**

当(17.1)中的常数项为一个可变项(某个 $t$ 的函数)代替时，仅仅会对特别积分产生影响。(为什么?)为求出新的特别积分，我

们仍可以应用待定系数法。在有关微分方程的内容中(15.6 节),待定系数法要求可变项及其逐阶导数一起(除乘积常数外)仅取有限个不同类型的表达式。将其应用于差分方程,这个要求应修改为"可变项及其逐次差分一起(除乘积常数外)仅取有限个不同类型的表达式"。我们通过一个具体的例子,先取形式为 $cm^t$ 的可变项,其中 $c$ 和 $m$ 为常数,来介绍这种方法。

**例 1** 求下列方程的特别积分

$$y_{t+2} + y_{t+1} - 3y_t = 7^t$$

这里,我们有 $c=1$ 和 $m=7$。首先,我们来确定可变项 $7^t$ 在逐次差分时是否会产生有限的表达式类型。按照差分法则($\Delta y_t = y_{t+1} - y_t$),$7^t$ 的一阶差分为

$$\Delta 7^t = 7^{t+1} - 7^t = (7-1)7^t = 6(7)^t$$

类似地,二阶差分,$\Delta^2(7^t)$,可以表示成

$$\Delta(\Delta 7^t) = \Delta 6(7^t) = 6(7)^{t+1} - 6(7)^t = 6(7-1)7^t = 36(7)^t$$

而且可以验证,所有各阶差分像一阶和二阶差分一样,均为 $7^t$ 的倍数。因为仅有一种表达式类型,所以我们可以试探 $y_t = B(7)^t$ 作为特别积分,其中 $B$ 为待定系数。

将试探解及$(t+1)$期、$(t+2)$期的对应形式带入给定的差分方程,得到

$$B(7)^{t+2} + B(7)^{t+1} - 3B(7)^t = 7^t \text{或} B(7^2+7-3)(7)^t = 7^t$$

因此,

$$B = \frac{1}{49+7-3} = \frac{1}{53}$$

可将特别积分写成

$$y_p = B(7)^t = \frac{1}{53}(7)^t$$

当然，它表示移动均衡。读者可以这样验证解的正确性：将其代入差分方程，从而看到产生恒等式 $7^t = 7^t$。

598　例 1 中所得到的结果很容易由可变项 $7^t$ 推广至 $cm^t$。根据经验，我们可以预期 $cm^t$ 所有各阶差分的表达式具有相同的形式：即 $Bm^t$，其中 $B$ 为某一乘积常数。因此，当给定差分方程

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = cm^t \tag{17.35}$$

我们可以试探解 $y_t = Bm^t$ 作为特别积分。运用试探解 $y_t = Bm^t$，这意味着 $y_{t+1} = Bm^{t+1}$，等等，我们可以将(17.35)重写成

$$Bm^{t+2} + a_1 Bm^{t+1} + a_2 Bm^t = cm^t$$

或

$$B(m^2 + a_1 m + a_2)m^t = cm^t$$

因此，试探解中的系数 $B$ 应当为

$$B = \frac{c}{m^2 + a_1 m + a_2}$$

而且所求的(17.35)的特别积分可以写成

$$y_p = Bm^t = \frac{c}{m^2 + a_1 m + a_2} m^t \quad (m^2 + a_1 m + a_2 \neq 0) \tag{17.36}$$

注意，$B$ 的分母不允许等于零。如果它恰好为零，[1] 那么，我们必须使用试探解 $y_t = Btm^t$，如果这个试探解也不行，则试探 $y_t = Bt^2 m^t$。

---

[1] 与 15.6 节中例 3 的情形类似，当常数 $m$ 恰好等于差分方程的特征根时，这种可能性就会成为现实。上述差分方程的特征根为满足方程 $b^2 + a_1 b + a_2 = 0$ 的 $b$ 值。如果一个特征根恰好等于 $m$，则必然可以导出 $m^2 + a_1 m + a_2 = 0$。

## 形式为 $ct^n$ 的可变项

现在我们考察形式为 $ct^n$ 的可变项，其中 $c$ 是任意常数，$n$ 为正整数。

**例 2** 求方程

$$y_{t+2} + 5y_{t+1} + 2y_t = t^2$$

的特别积分。求出 $t^2$($ct^n$ 当 $c = 1$ 和 $n = 2$ 时的特例)的前三阶差分如下：①

$$\Delta t^2 = (t+1)^2 - t^2 = 2t + 1$$

$$\Delta^2 t^2 = \Delta(\Delta t^2) = \Delta(2t+1) = \Delta 2t + \Delta 1$$

$$= 2(t+1) - 2t + 0 = 2 \quad [\Delta \text{常数} = 0]$$

$$\Delta^3 t^2 = \Delta(\Delta^2 t^2) = \Delta 2 = 0$$

因为进一步差分只能得到零，所以共有三种不同类型的表达式：$t^2$ 599 (源于可变项本身)，$t$ 和常数(源于可变项的各阶差分)。

因此，我们试用解

$$y_t = B_0 + B_1 t + B_2 t^2$$

作为特别积分，待定系数为 $B_0$，$B_1$ 和 $B_2$。注意，此解意味着

$$y_{t+1} = B_0 + B_1(t+1) + B_2(t+1)^2$$

$$= (B_0 + B_1 + B_2) + (B_1 + 2B_2)t + B_2 t^2$$

$$y_{t+2} = B_0 + B_1(t+2) + B_2(t+2)^2$$

---

① 这些结果可与 $t^2$ 的前三阶导数相比较：

$$\frac{d}{dt}t^2 = 2t \quad \frac{d^2}{dt^2}t^2 = 2 \text{ 和 } \frac{d^3}{dt^3}t^2 = 0$$

$$=(B_0+2B_1+4B_2)+(B_1+4B_2)t+B_2t^2$$

将其代入差分方程时，得到

$$(8B_0+7B_1+9B_2)+(8B_1+14B_2)t+8B_2t^2=t^2$$

使方程两边逐项相等，我们看到所求待定系数满足如下联立方程：

$$8B_0+7B_1+9B_2=0$$

$$8B_1+14B_2=0$$

$$8B_2=1$$

因此，待定系数值为 $B_0=13/256$, $B_1=-7/32$, $B_2=1/8$，给出特别积分

$$y_p=\frac{13}{256}-\frac{7}{32}t+\frac{1}{8}t^2$$

我们可以把处理可变项 $t^2$ 的方法推广至 $ct^2$ 的情况。在新的试探解中，显然应有一项 $B_nt^n$，与给定的可变项相对应。进而，因该项的逐阶差分产生不同的表达式 $t^{n-1}, t^{n-2}, \cdots, t, B_0$（常数），所以可变项 $ct^n$ 情况下的新的试探解应写成

$$y_t=B_0+B_1t+B_2t^2+\cdots+B_nt^n$$

但其余的步骤是完全相同的。

必须补充的是，这样的试探解也可能不成立。在这种情况下，我们还需要采用前面经常使用的技巧，即以 $t$ 的更高的幂数乘以原试探解。也就是说，我们可以试探 $y_t=t(B_0+B_1t+B_2t^2+\cdots+B_nt^n)$ 等等。

## 600 高阶线性差分方程

差分方程中的阶表示方程中差分的最高阶数，但它也表明所

包含的滞后期的最大数量。因此，一个具有常系数和常数项的 $n$ 阶线性差分方程一般可以写成：

$$y_{t+n}+a_1y_{t+n-1}+\cdots+a_{n-1}y_{t+1}+a_ny_t=c \qquad (17.37)$$

求此特别积分的方法与前面的方法并无本质区别。首先，我们仍可试用 $y_t=k$（静态瞬时均衡的情况）。如果此试探解不成立，则依次试探 $y_t=kt$，$y_t=kt^2$ 等等。

但在求余函数时，我们会遇到 $n$ 次多项式的特征方程

$$b^n+a_1b^{n-1}+\cdots+a_{n-1}b+a_n=0 \qquad (17.38)$$

这样会有 $n$ 个特征根 $b_i(i=1,2,\cdots,n)$，所有的特征根都将进入余函数：

$$y_c=\sum_{i=1}^{n}A_ib_i^t \qquad (17.39)$$

当然，这要假设所有根为不同实根。在存在重实根（比如 $b_1=b_2=b_3$）时，则和(17.39)中的前三项应修正为

$$A_1b_1^t+A_2tb_1^t+A_3t^2b_1^t \quad [参见(17.6)]$$

进而，如果存在一对共轭复根，比如 $b_{n-1}$，$b_n$，那么，和(17.39)中的最后两项应合并为表达式

$$R^t(A_{n-1}\cos\theta t+A_n\sin\theta t)$$

对其他任意一对共轭复根，也可以给出类似的表达式。但在存在两对**重根**的情况下，必须将其中的一对重根给予乘积因子 $tR^t$，而不是 $R^t$。

在求出 $y_p$ 和 $y_c$ 以后，将二者相加，便可以得到完全差分方程(17.37)的通解；即

$$y_t=y_p+y_c$$

但因为在此解中含有 $n$ 个任意常数，确定这个解至少需要 $n$ 个初

始条件。

601 **例3** 求三阶差分方程

$$y_{t+3}-\frac{7}{8}y_{t+2}+\frac{1}{8}y_{t+1}+\frac{1}{32}y_t=9$$

的通解。通过试探解 $y_t=k$,可以很容易求得 $y_p=32$。至于余函数,因为三次特征方程

$$b^3-\frac{7}{8}b^2+\frac{1}{8}b+\frac{1}{32}=0$$

可以分解成如下形式:

$$\left(b-\frac{1}{2}\right)\left(b-\frac{1}{2}\right)\left(b+\frac{1}{8}\right)=0$$

所以其根为 $b_1=b_2=1/2$ 和 $b_3=-1/8$。我们可以写出

$$y_c=A_1\left(\frac{1}{2}\right)^t+A_2t\left(\frac{1}{2}\right)^t+A_3\left(-\frac{1}{8}\right)^t$$

注意,上式第二项包含一个乘积因子 $t$,这是由于存在重根的缘故。这样,给定差分方程的通解不过是 $y_c$ 与 $y_p$ 的和而已。

在本例中,三个特征根的绝对值恰好都小于1,因此我们可以推断,得到的解表示收敛于静态均衡水平32的时间路径。

## 收敛性与舒尔(Schur)定理

当我们有一个难解的高阶差分方程时,不必努力求其实际定量解,我们仍可能确定有关时间路径的收敛性。读者可以回忆一下,当且仅当特征方程的每个根的绝对值均小于1时,时间路径才是收敛的。鉴于这一点,可直接应用如下定理(舒尔定理)[1]:

---

① 关于此定理及其历史的讨论,请参阅约翰·契普曼:《部门间货币流动与收入形成的理论》,119—120页。约翰·霍普金斯出版社,巴尔的摩,1951年。

$n$ 次多项式方程

$$a_0 b^n + a_1 b^{n-1} + \cdots + a_{n-1} b + a_n = 0$$

所有根的绝对值小于 1,当且仅当下列 $n$ 个行列式 602

$$\Delta_1 = \left|\begin{array}{c:c} a_0 & a_n \\ \hdashline a_n & a_0 \end{array}\right| \qquad \Delta_2 = \left|\begin{array}{cc:cc} a_0 & 0 & a_n & a_{n-1} \\ a_1 & a_0 & 0 & a_n \\ \hdashline a_n & 0 & a_0 & a_1 \\ a_{n-1} & a_n & 0 & a_0 \end{array}\right| \cdots$$

$$\Delta_n = \left|\begin{array}{cccc:cccc} a_0 & 0 & \cdots & 0 & a_n & a_{n-1} & \cdots & a_1 \\ a_1 & a_0 & \cdots & 0 & 0 & a_n & \cdots & a_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n-1} & a_{n-2} & \cdots & a_0 & 0 & 0 & \cdots & a_n \\ \hdashline a_n & 0 & \cdots & 0 & a_0 & a_1 & \cdots & a_{n-1} \\ a_{n-1} & a_n & \cdots & 0 & 0 & a_0 & \cdots & a_{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_1 & a_2 & \cdots & a_n & 0 & 0 & \cdots & a_0 \end{array}\right|$$

均为正。

注意,因为定理中的条件是在"当且仅当"的基础上给出的,所以它是充分必要条件。因此,完全差分方程中的舒尔定理与前面在微分方程中所介绍的罗斯定理是完全一致的。

这些行列式的构建并不复杂。利用虚线把行列式分成四个区域,就可以对其进行解释。$k$ 阶行列式 $\Delta_k$ 的每个区域,总是由 $k \times k$ 子行列式构成。左上方区域的对角元素均为 $a_0$,对角线上方的元素均为零,对角元素下方,每列中各元素的下标依次逐步增

大。转置左上方区域中的元素，就可以得到右下方区域。现在再考察右上方区域，对角元素仅由 $a_n$ 构成，对角线以下的元素均为零，当我们自对角元素沿每列向上时，各系数的下标依次变小。将此区域中的元素转置，便得到左下方区域。

可以直接应用此定理。因为特征方程的系数与出现在原差分方程左边的系数是相同的，所以，可直接将其引入上面所提及的行列式。

注意，在这里，我们总有 $a_0=1$。

**例 4** 方程

$$y_{t+2}+3y_{t+1}+2y_t=12$$

的时间路径收敛吗？这里我们有 $n=2$，系数为 $a_0=1$，$a_1=3$，$a_2=2$。因此有

$$\Delta_1=\begin{vmatrix} a_0 & a_2 \\ a_2 & a_0 \end{vmatrix}=\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix}=-3<0$$

603 因为它已违背了收敛性的条件，所以不必再推导 $\Delta_2$。

实际上，此差分方程的特征根很容易求得为 $b_1$，$b_2=-1$，$-2$，它确实意味着发散的时间路径。

**例 5** 运用舒尔定理，验证 $y_{t+2}+\frac{1}{6}y_{t+1}-\frac{1}{6}y_t=2$ 的时间路径是收敛的。这里的系数为 $a_0=1$，$a_1=1/6$，$a_2=-1/6(n=2)$。因此有

$$\Delta_1=\begin{vmatrix} a_0 & a_2 \\ a_2 & a_0 \end{vmatrix}=\begin{vmatrix} 1 & -\frac{1}{6} \\ -\frac{1}{6} & 1 \end{vmatrix}=\frac{35}{36}>0$$

$$\Delta_2 = \begin{vmatrix} a_0 & 0 & a_2 & a_1 \\ a_1 & a_0 & 0 & a_2 \\ a_2 & 0 & a_0 & a_1 \\ a_1 & a_2 & 0 & a_0 \end{vmatrix} = \begin{vmatrix} 1 & 0 & -\frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & 1 & 0 & -\frac{1}{6} \\ -\frac{1}{6} & 0 & 1 & \frac{1}{6} \\ \frac{1}{6} & -\frac{1}{6} & 0 & 1 \end{vmatrix} = \frac{1176}{1296} > 0$$

它们确实满足收敛性的充分必要条件。

## 练习 17.4

---

1 应用差分符号 Δ 的定义,求

($a$) $\Delta t$　　($b$) $\Delta^2 t$　　($c$) $\Delta t^3$

并将这些差分的结果与微分的结果相比较。

2 求下列方程的特别积分:

($a$) $y_{t+2} + 2y_{t+1} + y_t = 3^t$

($b$) $y_{t+2} - 5y_{t+1} - 6y_t = 2(6)^t$

($c$) $3y_{t+2} + 9y_t = 3(4)^t$

3 求下列方程的特别积分:

($a$) $y_{t+2} - 2y_{t+1} + 5y_t = t$

($b$) $y_{t+2} - 2y_{t+1} + 5y_t = 4 + 2t$

($c$) $y_{t+2} + 5y_{t+1} + 2y_t = 18 + 6t + 8t^2$

4 当可变项取 $m^t + t^n$ 形式时,你能断定试探解的形式为 $B(m)^t + (B_0 + B_1 t + \cdots + B_n t^n)$吗? 为什么?

5 求下列方程的特征根和余函数:

($a$) $y_{t+3} - \frac{1}{2}y_{t+2} - y_{t+1} + \frac{1}{2}y_t = 0$

($b$) $y_{t+3} - 2y_{t+2} + \frac{5}{4}y_{t+1} - \frac{1}{4}y_t = 1$

$\left[\text{提示:在每个特征方程提取公因子}\left(b-\frac{1}{2}\right)\text{。}\right]$

604 6 用舒尔定理检验下列差分方程的解的收敛性:

(a) $y_{t+2}+\frac{1}{2}y_{t+1}-\frac{1}{2}y_t=3$

(b) $y_{t+2}-\frac{1}{9}y_t=1$

7 在三阶差分方程

$$y_{t+3}+a_1y_{t+2}+a_2y_{t+1}+a_3y_t=c$$

情况下,什么是舒尔定理所要求的行列式的确切形式?

# 第18章　联立微分方程与差分方程

605

迄今为止,我们对经济动态学的讨论还只限定于对单个动态(微分或差分)方程的分析。在本章,我们将介绍分析联立动态方程组的方法。因为解联立方程组需要同时处理多个变量,所以读者可能估计会遇到更为复杂的情况。但毫无疑问,我们已经掌握的关于单一动态方程的方法完全可以拓展到联立动态方程组中。例如,动态方程组的解仍然由一组特别积分(各变量的瞬时均衡值)和余函数(与均衡的偏离)构成。余函数仍以简化方程,即方程组中的齐次方程为基础。方程组的动态稳定性仍取决于余函数中的特征根的符号(如果是微分方程组)或者绝对值(如果是差分方程组)。因此,动态方程组的问题只不过比单个动态方程组稍微复杂一点而已。

## 18.1　动态方程组的起源

有两种产生动态方程组的一般方式。一种可能源于一组给定的相互作用变化模式。另一种可能源于某一给定的变化模式,如 606 果这种变化模式包含一个二阶或更高阶动态方程的话。

## 相互作用的变化模式

最常见的一组给定的相互作用变化模式是多部门模型。模型中的每个部门,正如动态方程所描述的那样,至少与另外一个部门紧密联系。例如,动态的投入产出模型可能包含几个产业,一个产业产出的变化会对其它产业产生动态的影响。因此,它便构成了一个动态方程组。类似地,一个动态均衡市场模型可能包括几种在价格调整方面相互联系的商品。因此,它也构成了一个动态方程组。

然而,也可能在单一部门模型中找到相互作用变化模式。在此模型中的各个变量不再表示不同的部门或不同的商品,而是表示一个经济体的不同方面。不过,在其动态行为中,这些不同的方面能够互相影响,从而产生一种相互作用的网络。[①] 这种情况的具体例子实际上在17章中我们已经遇到过。在通货膨胀——失业模型中,预期的通货膨胀$\pi$所遵从的变化模式(17.19)不仅依赖于$\pi$,而且也依赖于失业率$U$(通过实际通货膨胀率$p$)。反过来看,$U$的变化模式(17.20)则取决于$\pi$(也通过$p$)。因此,$\pi$和$U$的动态过程必定为联合决定的。因此,回过头来看,通货膨胀失业模型可以被视为联立方程动态模型。而且这样还可以免去代入和消元以将方程化为单个变量的方程的冗长过程。在后面18.4节中,我们还将再研究这一模型,不过这次是将其作为动态方程组。

---

① 注意,如果我们有两个关于变量$y_1$和$y_2$的动态方程,使得$y_1$的变化模式完全取决于$y_1$本身,$y_2$的变化模式仅取决于$y_2$自身,那么,我们实际上并没有得到一个联立方程组,所得到的只不过是两个独立的动态方程,每一个均可独立进行分析,没有必要将其“联立”起来。

同时，同样一个模型既可以按照单一方程来分析，也可以作为方程组来分析，为我们讨论得到动态方程组的第二种方法提供一个自然线索。

## 高阶动态方程的变换

假定我们有一个 $n$ 阶单变量微分(或差分)方程。那么，正如将要证明的那样，总可能将此方程变换为在数学上等价的包含 $n$ 个变量的 $n$ 个联立一阶微分(或差分)方程组。特别地，一个二阶差分方程可重写为两个双变量联立一阶差分方程组。[①] 因此，即使我们恰好仅由一个(高阶)动态方程起步，我们还是可以通过数学变换的方法推导出一个动态方程组。顺便指出，这一事实具有非常重要的意义：在后面对动态方程组的讨论中，我们只需关注一阶方程组。因为若存在高阶方程，我们总可以首先将其变换为一组一阶方程。这会导致方程组中方程的数量增加，但却可把阶数降至最低。 607

为介绍变换方法，我们考察一个差分方程：

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = c \tag{18.1}$$

若我们创造一个新的人工变量，定义为

$$x_t \equiv y_{t+1} \qquad (\text{意味着 } x_{t+1} \equiv y_{t+2})$$

则我们可以用两个一阶联立(一期时滞)方程将原二阶方程表示如下：

$$\begin{aligned} x_{t+1} + a_1 x_t + a_2 y_t &= c \\ y_{t+1} - x_t &= 0 \end{aligned} \tag{18.1'}$$

① 反之，两个双变量一阶微分(或差分)方程组也可以化为一个单变量二阶方程。

容易看出，只要定义变量 $x_t$ 的第二个方程得到满足，那么，第一个方程与原给定方程是一致的。通过类似的方法并运用更多的人工变量，我们可以将一个更高阶数的方程变换为一个等价的一阶联立方程组。例如，读者可以验证，三阶方程

$$y_{t+3} + y_{t+2} - 3y_{t+1} + 2y_t = 0 \tag{18.2}$$

可以表示成

$$\begin{aligned} w_{t+1} \quad + w_t - 3x_t + 2y_t &= 0 \\ x_{t+1} \quad - w_t \qquad\qquad &= 0 \\ y_{t+1} \qquad\quad - x_t \qquad &= 0 \end{aligned} \tag{18.2'}$$

其中 $x_t \equiv y_{t+1}$（从而 $x_{t+1} \equiv y_{t+2}$），$w_t \equiv x_{t+1}$（从而 $w_{t+1} \equiv x_{t+2} \equiv y_{t+3}$）。

通过完全类似的步骤，我们还可以将一个几阶微分方程变换为一个由 $n$ 个一阶方程组成的方程组。例如，给定二阶微分方程

$$y''(t) + a_1 y'(t) + a_2 y(t) = 0 \tag{18.3}$$

我们可以引入一个新的变量 $x(t)$，定义为

$x(t) \equiv y'(t)$ ［意味着 $x'(t) \equiv y''(t)$］

则(18.3)可以重写为由两个一阶方程组成的方程组

$$\begin{aligned} x'(t) \quad + a_1 x(t) + a_2 y(t) &= 0 \\ y'(t) \quad - \quad x(t) \qquad\qquad &= 0 \end{aligned} \tag{18.3'}$$

读者可能会注意到，同(18.1′)中的第二个方程一样，(18.3′)中的第二个方程起着定义新引入的变量 $x$ 的作用。还可以用同样的步骤来变换更高阶的微分方程，唯一需要修正的是，在变换过程中需要相应地引进更多的新变量。

608

## 18.2 解联立动态方程

解联立微分方程和联立差分方程的方法是极为相似的，所以我们在本节将一起对其进行讨论。为此，我们将讨论仅限定在具有常系数的线性方程。

### 联立差分方程

假设我们有如下线性差分方程组：

$$\begin{aligned} x_{t+1} + 6x_t + 9y_t &= 4 \\ y_{t+1} - x_t &= 0 \end{aligned} \tag{18.4}$$

我们如何求 $x$ 和 $y$ 的时间路径，从而使得方程组中的两个方程均能得到满足呢？实质上，我们的任务就是求出特别积分和余函数，将二者相加便可以得到所要求的两个变量的时间路径。

因为特别积分表示瞬时均衡值，所以我们以 $\bar{x}$ 和 $\bar{y}$ 表示之。同前面一样，首先试探常数解，即 $x_{t+1}=x_t=\bar{x}$ 和 $y_{t+1}=y_t=\bar{y}$ 是明智的。在本例中，这种作法确实有效，因为将那些试探解代入(18.4)，得到

$$\left.\begin{aligned} 7\bar{x} + 9\bar{y} &= 4 \\ -\bar{x} + \bar{y} &= 0 \end{aligned}\right\} \Rightarrow \bar{x} = \bar{y} = \frac{1}{4} \tag{18.5}$$

（但在这种常数解不成立时，我们则需试探形式为 $x_t = k_1 t$，$y_t = k_2 t$ 的解，等等。）

根据以往的经验，对于余函数，我们应采用形式为

$$x_t = mb^t \qquad 和 \qquad y_t = nb^t \tag{18.6}$$

的试探解，其中 $m$ 和 $n$ 为任意常数，底 $b$ 代表特征根。则它自然意味着

$$x_{t+1}=mb^{t+1} \quad 和 \quad y_{t+1}=nb^{t+1} \tag{18.7}$$

注意，为使问题简化，对两个变量我们使用相同的底 $b\neq0$，但允许其系数有所不同。我们的目的就是求出 $b$，$m$ 和 $n$ 的值，使试探解(18.6)满足(18.4)的简化(齐次)形式。

609　将这些试探解代入(18.4)的简化形式，并消去共同因子 $b^t\neq0$，我们可以得到两个方程

$$\begin{aligned}(b+6)m+9n&=0\\ -m+bn&=0\end{aligned} \tag{18.8}$$

可以将其视为两变量 $m$ 和 $n$ 的线性齐次方程组——如果我们愿意将 $b$ 暂时视为参数的话。因为方程组(18.8)是齐次的，所以如果其系数矩阵是非奇异的(参见5.5节表5.1)，那么它将仅能产生试探解 $m=n=0$。在此情况下，(18.6)中的余函数均将等于零，表明 $x$ 和 $y$ 与其瞬时均衡值 $\bar{x}$ 和 $\bar{y}$ 不会出现偏离。因为这是一个没有什么意义的特例，我们应通过要求方程组的系数矩阵为**奇异的**，以设法排除那种试探解。即，我们将要求矩阵的行列式为零：

$$\begin{vmatrix}b+6 & 9\\ -1 & b\end{vmatrix}=b^2+6b+9=0 \tag{18.9}$$

由此二次方程，我们发现 $b(=b_1=b_2)=-3$ 是使(18.8)中的 $m$ 和 $n$ 均免于同时为零的唯一值，因此我们将仅使用这个 $b$ 值。方程(18.9)被称作给定联立差分方程组的**特征方程**，其根被称作给定联立差分方程组的**特征根**。

一旦我们得到 $b$ 的具体值，(18.8)便会给出对应的 $m$ 和 $n$ 的解值。然而，作为一个齐次方程组，实际上将会出现无穷多个($m$，

$n$)的解，可用方程 $m=kn$ 表示之，其中 $k$ 为常数。事实上，对于每一个根 $b_i$，一般总会有不同解的方程 $m_i=k_in$。但在本例 $b=-3$ 的重根的情况下，我们仅会得到单一方程 $m=-3n$。

如果我们现在令 $n=A_3$($A_3$ 为任意常数)，从而 $m=-3A_3$，则回忆一下 $b=-3$，试探解(18.6)将会取具体形式

$$x_t=-3A_3(-3)^t \quad 和 \quad y_t=A_3(-3)^t$$

且余函数将会写成

$$x_c=-3A_3(-3)^t-3A_4(t+1)(-3)^t \qquad [由(17.6)]$$

$$y_c=A_3(-3)^t+A_4t(-3)^t \tag{18.10}$$

其中为满足(18.4)中的条件 $x_t=y_{t+1}$，使用了乘积项($t+1$)以代替 $x_c$ 的第二项中的 $t$。则将(18.5)中的特别积分与刚求得的余函数合并便可得到通解。

上述解的一个显著特征是，因为两个时间路径中具有一致的 $b^t$ 表达式，所以它们必然同时收敛或者发散。在一个具有动态的相互作用的变量的模型中，这具有特殊的意义，因为除非在方程组中不存在任何动态运动，否则不可能获得一般瞬时均衡。本例具 610 有重根 $b=-3$，$x$ 和 $y$ 的时间路径均表现为放大振荡。

## 矩阵符号

为揭示单个方程与方程组解法之间的相似性，我们前面进行描述时没有使用矩阵符号。现在我们来看如何应用这些符号。虽然把矩阵符号应用到仅含两个方程的简单方程组似乎没有多大意义，但由于存在将其推广至含 $n$ 个方程的情况的可能性，所以这是一个值得做的练习。

首先,已知方程(18.4)可以表示成

$$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}x_{t+1}\\y_{t+1}\end{bmatrix}+\begin{bmatrix}6 & 9\\-1 & 0\end{bmatrix}\begin{bmatrix}x_t\\y_t\end{bmatrix}=\begin{bmatrix}4\\0\end{bmatrix} \tag{18.4'}$$

或更简洁地

$$Iu + Kv = d \tag{18.4''}$$

其中 $I$ 为 $2\times2$ 单位矩阵;$K$ 为 $x_t$ 和 $y_t$ 项的 $2\times2$ 系数矩阵;$u$、$v$ 和 $d$ 为定义如下的列向量:[①]

$$u=\begin{bmatrix}x_{t+1}\\y_{t+1}\end{bmatrix}\quad v=\begin{bmatrix}x_t\\y_t\end{bmatrix}\quad d=\begin{bmatrix}4\\0\end{bmatrix}$$

读者可能发现一个令人迷惑的特征:因为我们已知道 $Iu=u$,那为什么不去掉 $I$ 呢? 原因是,虽然 $I$ 现在看起来是多余的,但在后面的运算中仍需使用单位矩阵,因此我们仍像在(18.4″)中那样保留它。

当我们试探常数解 $x_{t+1}=x_t=\bar{x}$ 和 $y_{t+1}=y_t=\bar{y}$ 作为特别积分时,实际上我们是在令 $u=v=\begin{bmatrix}\bar{x}\\\bar{y}\end{bmatrix}$;这将使(18.4″)简化为

$$(I+K)\begin{bmatrix}\bar{x}\\\bar{y}\end{bmatrix}=d$$

如果逆 $(I+K)^{-1}$ 存在,我们可以将这一特别积分表述为:

$$\begin{bmatrix}\bar{x}\\\bar{y}\end{bmatrix}=(I+K)^{-1}d \tag{18.5'}$$

当然这是一般的公式,因为只要 $(I+K)^{-1}$ 存在,它对任意矩阵 $K$

---

① 这里符号 $v$ 表示向量。不要将其与复数符号 $h\pm v_i$ 中的 $v$ 相混淆。复数符号中的 $v$ 表示标量。

和向量 $d$ 都成立。将其应用于我们的数字例子,我们有

$$(I+K)^{-1}d=\begin{bmatrix}7 & 9\\ -1 & 1\end{bmatrix}^{-1}\begin{bmatrix}4\\ 0\end{bmatrix}=\begin{bmatrix}\frac{1}{16} & -\frac{9}{16}\\ \frac{1}{16} & \frac{7}{16}\end{bmatrix}\begin{bmatrix}4\\ \\ 0\end{bmatrix}=\begin{bmatrix}\frac{1}{4}\\ \frac{1}{4}\end{bmatrix}$$

因此,$\bar{x}=\bar{y}=1/4$,它与(18.5)一致。

转到余函数,我们看到试探解(18.6)和(18.7)给出 $u$ 和 $v$ 向 611 量的具体形式

$$u=\begin{bmatrix}mb^{t+1}\\ nb^{t+1}\end{bmatrix}=\begin{bmatrix}m\\ n\end{bmatrix}b^{t-1} \quad \text{和} \quad v=\begin{bmatrix}mb^{t}\\ nb^{t}\end{bmatrix}=\begin{bmatrix}m\\ n\end{bmatrix}b^{t}$$

当将其代入简化方程 $Iu+Kv=0$ 时,这些试探解将后者变成

$$I\begin{bmatrix}m\\ n\end{bmatrix}b^{t+1}+K\begin{bmatrix}m\\ n\end{bmatrix}b^{t}=0$$

或者以 $b^{-t}$(一个标量)通乘并提取公因子后,得到

$$(bI+K)\begin{bmatrix}m\\ n\end{bmatrix}=0 \qquad (18.8')$$

其中 0 为零向量。正是由此齐次方程组,我们要求出将在试探解中使用的 $b,m,n$ 的合适的值,从而使试探解明确下来。

为避免 $m$ 和 $n$ 的平凡解,需要

$$|bI+K|=0 \qquad (18.9')$$

这是将会给出特征根 $b_i$ 的特征方程。读者可以验证,如果我们将

$$bI=\begin{bmatrix}b & 0\\ 0 & b\end{bmatrix} \quad \text{和} \quad K=\begin{bmatrix}6 & 9\\ -1 & 0\end{bmatrix}$$

代入方程,结果便是能够产生重根 $b=-3$ 的方程(18.9)。

一般而言,每个根 $b_i$ 将由(18.8′)无数 $m$ 和 $n$ 解值中的特定的一组给出,$m$ 和 $n$ 的解值由方程 $m_i=k_in_i$ 联系起来。因此,对每个 $b_i$ 值可能写出

$$n_i = A_i \quad 和 \quad m_i = k_i A_i$$

其中 $A_i$ 为有待于确定的任意常数。将其代入试探解，这些 $n_i$ 和 $m_i$ 的表达式及 $b_i$ 值会引出一个特殊形式的余函数。若所有的根均为不同的实数，则我们可应用(17.5)并写成

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} \sum m_i b_i^t \\ \sum n_i b_i^t \end{bmatrix} = \begin{bmatrix} \sum k_i A_i b_i^t \\ \sum A_i b_i^t \end{bmatrix}$$

但在重根情况下，我们则必须应用(17.6)，结果对我们现在的数字例子，同(18.10)中的结果一样，余函数将含有像 $A_3 b^t$ 和 $A_4 + b^t$ 这样的项。最后，在复根情况下，应按其原型(17.10)那样写出余函数。

最后，为得到通解，只需求和

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} x_c \\ y_c \end{bmatrix} + \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}$$

则余下的工作只是确定任意常数 $A_i$。

612 将此方法推广至 $n$ 元方程组应是不言自明的。但当 $n$ 较大时，不容易定量解出特征方程($n$ 次多项式方程)。在此情况下，我们仍需借助于舒尔定理，来得出方程组中变量的时间路径的定性结论。我们记得，在试探解中，所有变量都被赋予同样的底 $b$，所以最后在余函数中有相同的 $b_i^t$ 式，且具有同样的收敛性。这样，只需一次应用舒尔定理，就可以确定方程组中每个变量的时间路径的敛散性。

## 联立微分方程

刚才介绍的解法也可以应用于一阶线性微分方程组。所需进行的唯一重要的修正是将试探解变成

$$x(t) = me^{rt} \quad 和 \quad y(t) = ne^{rt} \tag{18.11}$$

这意味着

$$x'(t) = rme^{rt} \quad 和 \quad y'(t) = rne^{rt} \tag{18.12}$$

与我们原来使用符号习惯相一致,特征根现在以 $r$ 而非以 $b$ 表示。

假设我们有如下方程组

$$\begin{aligned} x'(t) + 2y'(t) + 2x(t) + 5y(t) &= 77 \\ y'(t) + x(t) + 4y(t) &= 61 \end{aligned} \tag{18.13}$$

首先,我们将其改写为矩阵形式

$$Ju + Mv = g \tag{18.13'}$$

其中矩阵为

$$J = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \quad u = \begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix}$$

$$M = \begin{bmatrix} 2 & 5 \\ 1 & 4 \end{bmatrix} \quad v = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \quad g = \begin{bmatrix} 77 \\ 61 \end{bmatrix}$$

注意,鉴于(18.13)第一个方程中存在 $2y'(t)$ 项,我们必须使用矩阵 $J$ 代替(18.4″)中的单位矩阵 $I$。当然,若 $J$ 为非奇异的(从而 $J^{-1}$ 存在),则我们可以 $J^{-1}$ 左乘(18.13′)各项,使其在某种意义上正规化,以得到

$$J^{-1}Ju + J^{-1}Mv = J^{-1}g \quad 或 \quad Iu + Kv = d$$

$$(K \equiv J^{-1}M; d \equiv J^{-1}g) \tag{18.13''}$$

这个新的形式与(18.4″)完全一致,但必须记住向量 $u$ 和 $v$ 在两种情况下具有完全不同的意义。在后面的推导中,我们将始终使用(18.13′)中给出的公式 $Ju + Mv = g$。

为求出特别积分,我们试探常数解 $x(t) = \bar{x}$ 和 $y(t) =$ 613 $\bar{y}$——这意味着 $x'(t) = y'(t) = 0$。若这两个解成立,则向量 $u$

和 $v$ 将变成 $v=\begin{bmatrix}\bar{x}\\ \bar{y}\end{bmatrix}$；$u=\begin{bmatrix}0\\0\end{bmatrix}$，且(18.13′)会化简为 $Mv=g$。因此，$\bar{x}$ 和 $\bar{y}$ 的解可以写成

$$\begin{bmatrix}\bar{x}\\ \bar{y}\end{bmatrix}=v=M^{-1}g \tag{18.14}$$

读者可将其与(18.5′)相比较。用数字表示，本题得到如下特别积分

$$\begin{bmatrix}\bar{x}\\ \bar{y}\end{bmatrix}=\begin{bmatrix}2 & 5\\1 & 4\end{bmatrix}^{-1}\begin{bmatrix}77\\61\end{bmatrix}=\begin{bmatrix}\frac{4}{3} & -\frac{5}{3}\\ -\frac{1}{3} & \frac{2}{3}\end{bmatrix}\begin{bmatrix}77\\61\end{bmatrix}=\begin{bmatrix}1\\15\end{bmatrix}$$

其次，我们来求余函数。运用(18.11)和(18.12)中所提示的试探解，向量 $u$ 和 $v$ 变成

$$u=\begin{bmatrix}m\\n\end{bmatrix}re^{rt} \quad 和 \quad v=\begin{bmatrix}m\\n\end{bmatrix}e^{rt}$$

将其代入简化方程

$$Ju+Mv=0$$

产生结果

$$J\begin{bmatrix}m\\n\end{bmatrix}re^{rt}+M\begin{bmatrix}m\\n\end{bmatrix}e^{rt}=0$$

或在以标量 $e^{-rt}$ 通乘并提取公因子后，

$$(rJ+M)\begin{bmatrix}m\\n\end{bmatrix}=0 \tag{18.15}$$

读者可将其与(18.8′)相比较。因为我们的目的是求出 $m$ 和 $n$ 的非平凡解(因而我们的试探解也将是非平凡的)，所以必须

$$|rJ+M|=0 \tag{18.16}$$

最后这个方程(已知方程组的特征方程)将会产生我们要求的根 $r_i$,则我们可以求出相应的(非平凡的)$m_i$ 和 $n_i$ 值。

在本例中,特征方程为

$$|rJ + M| = \begin{vmatrix} r+2 & 2r+5 \\ 1 & r+4 \end{vmatrix} = r^2 + 4r + 3 = 0 \tag{18.16'}$$

其根为 $r_1 = -1, r_2 = -3$。将其代入(18.15),得到

$$\begin{bmatrix} 1 & 3 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} m_1 \\ n_1 \end{bmatrix} = 0 \quad (\text{对于 } r_1 = -1)$$

$$\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} m_2 \\ n_2 \end{bmatrix} = 0 \quad (\text{对于 } r_2 = -3)$$

由此得到 $m_1 = -3n_1, m_2 = -n_2$,它们也可表示成

614

$$m_1 = 3A_1 \quad \text{和} \quad m_2 = A_2$$

$$n_1 = -A_1 \qquad n_2 = -A_2$$

现在 $r_i, m_i, n_i$ 均已求出,则余函数可以写成如下指数式的线性组合:

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} \sum m_i e^{r_i t} \\ \sum n_i e^{r_i t} \end{bmatrix} \quad [\text{不同的实根}]$$

且通解的形式为

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} x_c \\ y_c \end{bmatrix} + \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}$$

在本例中,解为

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} 3A_1 e^{-t} + A_2 e^{-3t} + 1 \\ -A_1 e^{-t} - A_2 e^{-3t} + 15 \end{bmatrix}$$

进而,如果已知初始条件 $x(0) = 6, y(0) = 12$,则可求得任意常数

为 $A_1=1, A_2=2$。可用它们来确定上面的解。

我们可以再一次观测到，因 $e^{r_it}$ 式为时间路径 $x(t)$ 和 $y(t)$ 所共有，所以 $x(t)$ 和 $y(t)$ 必然或者均为收敛或者均为发散。本例中的根为 $-1$ 和 $-3$，所以两个时间路径分别收敛于其均衡值，即 $\bar{x}=1$ 和 $\bar{y}=15$。

尽管本例仅是一个由两个方程组成的方程组，但处理此方程组的方法无疑可以推广到含 n 个方程的一般方程组。虽然 $n$ 很大时，很难求出定量解，但一旦求出特征方程，便可以利用罗斯定理进行定性分析。

### 对特征方程的进一步说明

“特征方程”这个术语我们现在已在三种不同场合中遇到过：在 11.3 节，我们谈到矩阵的特征方程；在 15.1 节和 17.1 节，这一术语又被应用于单个的线性微分方程和差分方程；现在在本节，我们已介绍了线性差分或微分方程组的特征方程。这三者之间存在某种联系吗？

这三者之间确有联系，而且联系密切。首先，给定的单个方程和等价的方程组〔以方程(18.1)和方程组(18.1′)，或方程(18.3)和方程组(18.3′)为例〕的特征方程必然是一致的。为便于说明，考察差分方程(18.1)，$y_{t+2}+a_1y_{t+1}+a_2y_t=c$。我们前面已学615过，直接将其常系数移植到二次方程中便可写出特征方程：

$$b^2+a_1b+a_2=0$$

那么，等价的(18.1′)又如何呢？如在(18.4″)中那样，取方程组的形式为 $Iu+Kv=d$，我们有矩阵 $K=\begin{bmatrix} a_1 & a_2 \\ -1 & 0 \end{bmatrix}$。从而特征方程

为

$$|bI+K| = \begin{vmatrix} b+a_1 & a_2 \\ -1 & b \end{vmatrix}$$

$$= b^2 + a_1 b + a_2 = 0 \quad [由(18.9')] \qquad (18.17)$$

它与单个方程的特征方程是完全一致的。自然，在微分方程框架中，同样的结果也是成立的，唯一的差别是，根据习惯在后一种框架中我们要以符号 $r$ 代替符号 $b$。

将差分(或微分)方程组的特征方程与特殊方阵(称之为 $D$)的特征方程联系起来也是可能的。参照(11.14)中的定义，但在差分方程框架中使用符号 $b$(而非 $r$)，我们可写出矩阵 $D$ 的特征方程如下：

$$|D - bI| = 0 \qquad (18.18)$$

一般而言，若以 $-1$ 乘行列式 $|D-bI|$ 中的每一元素，如果矩阵 $D$ 包含偶数行(或列)，则行列式的值不变；如果矩阵 $D$ 包含奇数行或列，则行列式改变符号。但在这里，因为 $|D-bI|$ 等于零，以 $-1$ 乘以每个元素没有什么意义，与矩阵 $D$ 的维数无关。但以 $-1$ 乘行列式 $|D-bI|$ 的每一元素相当于在取矩阵 $(D-bI)$ 的行列式以前乘以 $-1$(参见 5.3 节例 6)。因此，(18.18)可以重写为

$$|bI - D| = 0 \qquad (18.18')$$

使之与(18.17)相等时，若我们选择矩阵 $D=-K$，则其特征方程将与方程组(18.1′)的特征方程恒等。此矩阵 $-K$ 具有特殊意义：若我们取方程组的简化型 $Iu+Kv=0$，并将其表示为 $Iu=-Kv$，或简单地 $u=-Kv$，则我们看到 $-K$ 在这个特殊方程中是一个可将向量 $v=\begin{bmatrix} x_t \\ y_t \end{bmatrix}$ 转化为向量 $u=\begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix}$ 的矩阵。

同样的推理也适用于微分方程组(18.3′)。但是在方程组(18.13′)$Ju+Mv=g$ 情况下〔不同于方程组(18.3′),其第一项为 $Ju$ 而非 $Iu$〕,我们不应取 $D=-M$。相反,我们必须首先将方程 $Ju+Mv=g$ 正规化为(18.13″)的形式,然后再取 $D=-K=-J^{-1}M$。

总之,给定(1)单个差分或微分方程,及(2)一个等价的方程组,由此方程组我们也可得到(3)一个适当的矩阵 $D$,616 如果我们设法求出三种情况的特征方程,那么,其结果必然是完全一致的。

## 练习 18.2

---

1　验证差分方程组(18.4)与单个方程 $y_{t+2}+6y_{t+1}+9y_t=4$ 等价,此方程已在 17.1 节例 4 中解出。如何比较用两种不同方法所得到的解?

2　证明差分方程(18.2)的特征方程与等价方程组(18.2′)的特征方程是完全一致的。

3　解下列差分方程组:

(a)
$$x_{t+1}+x_t+2y_t=24$$
$$y_{t+1}+2x_t-2y_t=9\quad(\text{其中 } x_0=10 \text{ 和 } y_0=9)$$

(b)
$$x_{t+1}-x_t-\frac{1}{3}y_t=-1$$
$$x_{t+1}+y_{t+1}-\frac{1}{6}y_t=8\frac{1}{2}\quad(\text{其中 } x_0=5 \text{ 和 } y_0=4)$$

4　解下列差分方程组:

(a)
$$x'(t)-x(t)-12y(t)=-60$$
$$y'(t)+x(t)+6y(t)=36\quad[\text{其中 } x(0)=13 \text{ 和 } y(0)=4]$$

(b)
$$x'(t)-2x(t)+3y(t)=10$$
$$y'(t)-x(t)+2y(t)=9\quad[\text{其中 } x(0)=8 \text{ 和 } y(0)=5]$$

5　以微分方程组(18.13)为基础，求与该方程组有相同特征方程的矩阵 $D$。验证这两个特征方程确实相同。

---

## 18.3　动态投入产出模型

投入产出分析首先遇到的问题是：每个产业究竟生产多少产品，才能恰好满足所有产业的投入需求，以及(开放系统)的最终需求？这是一个静态的解联立方程组以求所有产业的均衡产出水平的问题。当把某些额外的经济条件纳入模型时，投入产出系统会表现出动态特征，从而产生上节讨论过的差分或微分方程组。

这里需要考察三种动态条件。但为使表述简洁，我们将仅说明两个部门的开放系统。然而因为我们将使用矩阵符号，所以将其推广到 $n$ 产业部门的情况并不会产生困难，只需适当改变矩阵的维数便可以做到这一点。为便于推广，使用符号 $x_{1,t}$ 和 $x_{2,t}$ 而非 $x_t$ 和 $y_t$ 要更明智一些，因为当需要时，我们便可以将符号推广至 $x_{n,t}$。读者可以回顾一下，在投入—产出分析的有关内容中，$x_i$ 617 表示第 $i$ 个产业部门的产出(以美元度量)；而现在新的下标 $t$ 则把时间因素加到相应的产业部门。投入系数符号 $a_{ij}$ 仍表示生产一美元的第 $j$ 种商品所需要的第 $i$ 种商品的美元数，$d_i$ 仍表示第 $i$ 种商品的最终需求。

### 生产中的时滞

在静态的两产业开放系统中，产业 I 的产业应确定在如下需求水平：

$$x_1 = a_{11}x_1 + a_{12}x_2 + d_1$$

现在假设生产时滞为一个时期，因而时期 $t$ 的需要量不是决定现期产出，而是 $(t+1)$ 期的产出。为描述这种新的情况，我们必须将上述方程修正为这种形式

$$x_{1,t+1} = a_{11}x_{1,t} + a_{12}x_{2,t} + d_{1,t} \tag{18.19}$$

类似地，我们可以写出产业 II 的方程

$$x_{2,t+1} = a_{21}x_{1,t} + a_{22}x_{2,t} + d_{2,t} \tag{18.19'}$$

这样我们现在有了一个联立差分方程组。它构成了投入产出模型的动态形式。

用矩阵符号，方程组由下列方程构成：

$$x_{t+1} - Ax_t = d_t \tag{18.20}$$

其中

$$x_{t+1} = \begin{bmatrix} x_{1,t+1} \\ x_{2,t+1} \end{bmatrix} \quad x_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix}$$

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad d_t = \begin{bmatrix} d_{1,t} \\ d_{2,t} \end{bmatrix}$$

显然，(18.20)与(18.4″)形式一致，但有两点例外。首先，不同于向量 $\boldsymbol{u}$，向量 $x_{t+1}$ 并没有单位矩阵 $I$ 作为其“系数”，但正如前面所解释的那样，这并无分析上的差别。其次，也是更为实质的一点是，具有时间下标的向量 $d_t$ 意味着最终需求向量被视为时间的函数。如果此函数不为常数，则求特别积分的方法需要修正，尽管余函数不受影响。下面的例子可说明修正的步骤。

**例 1**　已知指数式最终需求向量

$$d_t = \begin{bmatrix} \delta^t \\ \delta^t \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \delta^t \quad (\delta = \text{一个正标量})$$

求动态投入一产出模型(18.20)的特别积分。根据 17.4 节介绍的待定系数法，我们将试探形式为 $x_{1,t} = \beta_1 \delta^t$ 和 $x_{2,t} = \beta_2 \delta^t$ 的解，其

中 $\beta_1$ 和 $\beta_2$ 为待定系数。即,我们应试探 618

$$x_t = \begin{bmatrix} \beta_1\delta^t \\ \beta_2\delta^t \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}\delta^t \tag{18.21}$$

它意味着①

$$x_{t+1} = \begin{bmatrix} \beta_1\delta^{t+1} \\ \beta_2\delta^{t+1} \end{bmatrix} = \begin{bmatrix} \beta_1\delta \\ \beta_2\delta \end{bmatrix}\delta^t = \begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}\delta^t$$

如果所提出的试探解成立,那么(18.20)将变成

$$\begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}\delta^t - \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}\delta^t = \begin{bmatrix} 1 \\ 1 \end{bmatrix}\delta^t$$

或者消去标量公因子 $\delta^t \neq 0$,

$$\begin{bmatrix} \delta - a_{11} & -a_{12} \\ -a_{21} & \delta - a_{22} \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{18.22}$$

假定最左边的系数矩阵为非奇异的,我们可以根据克莱姆法则求出 $\beta_1$ 和 $\beta_2$ 为

$$\beta_1 = \frac{\delta - a_{22} + a_{12}}{\Delta} \quad 和 \quad \beta_2 = \frac{\delta - a_{11} + a_{21}}{\Delta} \tag{18.22'}$$

其中 $\Delta \equiv (\delta - a_{11})(\delta - a_{22}) - a_{12}a_{21}$。因为 $\beta_1$ 和 $\beta_2$ 现在完全以已知参数表示,我们仅需将其代入试探解(18.21),以得到特别积

---

① 读者应注意到向量 $\begin{bmatrix} \beta_1\delta \\ \beta_2\delta \end{bmatrix}$ 可以重写成好几种等价的形式:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}\delta \quad 或 \quad \delta\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \quad 或 \quad \delta\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

这里我们选择第三种形式,因为下一步需要把 $\begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix}$ 加到另一个 $2 \times 2$ 的矩阵中去。前两种可供选择的形式要求问题满足维数可相乘条件。

分的确定表达式。

这里讨论的最终需求向量的更一般的形式将在练习 18.3－1 中给出。

求(18.20)余函数的步骤与上一节所介绍的方法没有什么不同。因方程组的齐次形式为 $x_{t+1}-Ax_t=0$,所以特征方程应为

$$|bI-A|=\begin{vmatrix} b-a_{11} & -a_{12} \\ -a_{21} & b-a_{22} \end{vmatrix}=0 \quad [\text{参见}(18.9')]$$

由此我们可以求出特征根 $b_1$ 和 $b_2$,并继续求解过程剩下的步骤。

## 619 超额需求与产出调整

(18.20)中的模型表达式也可产生于不同的经济假设。考察这样一种情况:对每一产品的超额需求会引致产出增量等于超额需求。因为在时期 $t$ 对第一种产品的超额需求等于

$$\underbrace{a_{11}x_{1,t}+a_{12}x_{2,t}+d_{1,t}}_{(\text{需求的})}-\underbrace{x_{1,t}}_{(\text{供给的})}$$

必使产出调整(增量)$\Delta x_1$ 恰好等于该水平:

$$\Delta x_{1,t}(\equiv x_{1,t+1}-x_{1,t})=a_{11}x_{1,t}+a_{12}x_{2,t}+d_{1,t}-x_{1,t}$$

但若我们在方程两边同时加上 $x_{1,t}$,其结果将与(18.19)相同。类似地,对第二个产业部门,我们的产出调整假设将给出一个与(18.19′)相同的方程。

至此,我们一直在离散时间框架下考察投入—产出系统。为了比较,我们现在在连续时间模式下考察产出调整过程。

在连续时间框架下,要求使用符号 $x_i(t)$ 以代替 $x_{i,t}$,使用导数 $x'_i(t)$ 以代替 $\Delta x_{i,t}$。根据这些变化,产出调整假设可以表示成如下两个微分方程:

$$x'_1(t) = a_{11}x_1(t) + a_{12}x_2(t) + d_1(t) - x_1(t)$$
$$x'_2(t) = a_{21}x_1(t) + a_{22}x_2(t) + d_2(t) - x_2(t)$$

在任意时间 $t = t_0$,符号 $x_i(t_0)$表示在该时刻的单位时间的产出流量率,$d_i(t_0)$表示那一时刻每月的最终需求。因此每个方程等号右边的和表示在 $t = t_0$ 度量的每月的超额需求率。而左边的导数 $x_i'(t_0)$表示在 $t = t_0$ 的超额需求所引起的月产出调整率。只有在超额需求和产出调整均保持在现在的比率时,这个调整过程才会在一个月的时间内消除超额需求并导致均衡。实际上,超额需求像引致的产出调整一样,将随时间的变化而变化,从而产生"猫追老鼠"的游戏。方程组的解构成产出 $x_i$ 的时间路径,成为这场"追逐"的记录者。如果解是收敛的,那么,"猫"(产出调整)最终会逐渐追上"鼠"(超额需求)(当 $t \to \infty$时)。

经过适当重排,此微分方程组可按(18.13')的形式表示如下:

$$Ix' + (I - A)x = d \tag{18.23}$$

其中

$$x' = \begin{bmatrix} x'_1(t) \\ x'_2(t) \end{bmatrix} \quad x = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad d = \begin{bmatrix} d_1(t) \\ d_2(t) \end{bmatrix}$$

(符号"'"表示导数,而非转置)。余函数可以通过前面讨论的方法 620 求出。特别地,特征根可由如下方程求出

$$|rI + (I - A)| = \begin{vmatrix} r + 1 - a_{11} & -a_{12} \\ -a_{21} & r + 1 - a_{22} \end{vmatrix} = 0$$

[参见(18.16)]

至于特别积分，如果最终需求向量包含时间的非常数函数 $d_1(t)$、$d_2(t)$ 作为其元素，则求解方法必须作调整和修正。下面我们介述一个简单的例子。

**例 2** 已知最终需求向量

$$d = \begin{bmatrix} \lambda_1 e^{\rho t} \\ \lambda_2 e^{\rho t} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} e^{\rho t}$$

其中 $\lambda_i$ 和 $\rho$ 为常数，求动态模型(18.23)的特别积分。运用待定系数法，我们可以试探形式为 $x_i(t) = \beta_i e^{\rho t}$（当然，这意味着 $x_i'(t) = \rho\beta_i e^{\rho t}$）的解。用矩阵符号，可将其写成

$$x = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} e^{\rho t} \tag{18.24}$$

和 $$x' = \rho \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} e^{\rho t} = \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} e^{\rho t}$$ [参见前注]

代入(18.23)并消去公共(非零)标量乘子 $e^{\rho t}$，得到

$$\begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} 1 - a_{11} & -a_{12} \\ -a_{21} & 1 - a_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

或

$$\begin{bmatrix} \rho + 1 - a_{11} & -a_{12} \\ -a_{21} & \rho + 1 - a_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \tag{18.25}$$

如果最左边的矩阵为非奇异的，则可应用克莱姆法则并确定系数 $\beta_i$ 的值为

$$\beta_1 = \frac{\lambda_1(\rho + 1 - a_{22}) + \lambda_2 a_{12}}{\Delta}$$

$$\beta_2 = \frac{\lambda_2(\rho + 1 - a_{11}) + \lambda_1 a_{21}}{\Delta} \tag{18.25'}$$

其中 $\Delta \equiv (\rho+1-a_{11})(\rho+1-a_{22})-a_{12}a_{21}$。这样,待定系数确定以后,我们便可以将其值代入试探解(18.24)以得到所求的特别积分。

## 资本形成



产出动态投入产出系统的另一个经济考虑是资本形成,其中包括存货积累。

在静态的讨论中,我们仅考虑用于满足现实需求的每种产品的产出水平,对存货积累或资本形成的需要或者被忽视了,或者被归入最终需求向量。为揭开资本形成的奥秘,我们现在考察投入系数矩阵 $A=[a_{ij}]$和资本系数矩阵

$$C=[c_{ij}]=\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

其中 $c_{ij}$表示第 $j$ 个产业部门由于产出增加 1 美元而需要的第 $i$ 种商品的美元数;这种对第 $i$ 种商品的需要是作为新的资本(或者是设备,或者是存货,视商品的性质而定)。例如,如果软饮料产业(第 $j$ 个产业)产出增加 1 美元,导致装瓶设备第 $i$ 种产品增加 2 美元,则 $c_{ij}=2$。因此,这种资本系数表示某种边际资本产出比,此比率仅限定为一类资本(第 $i$ 种商品)。像投入系数 $a_{ij}$一样,假定资本系数也是固定的。对于一个经济体,它生产的每一产品的产量不仅要满足投入需求加最终需求,而且要满足资本形成对它的需求。

如果时间是连续的,则产出增量以导数 $x_i'(t)$表示;这样每个产业的产出应确定为

$$x_1(t)=a_{11}x_1(t)+a_{12}x_2(t)+c_{11}x_1'(t)+c_{12}x_2'(t)+d_1(t)$$

$$x_2(t) = \underbrace{a_{21}x_1(t) + a_{22}x_2(t)}_{\text{投入需求}} + \underbrace{c_{21}x_1'(t) + c_{22}x_2'(t)}_{\text{资本需求}} + \underbrace{d_2(t)}_{\text{最终需求}}$$

用矩阵符号，可以表示成方程

$$Ix = Ax + Cx' + d$$

或

$$Cx' + (A - I)x = -d \tag{18.26}$$

如果时间是**离散的**，$t$ 期的资本需求则会建立在产出增量 $x_{i,t} - x_{i,t-1}(\equiv \Delta_{i,t-1})$ 的基础之上。产出水平应确定为

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix}}_{\text{投入需求}}$$

$$+ \underbrace{\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} x_{1,t} - x_{1,t-1} \\ x_{2,t} - x_{2,t-1} \end{bmatrix}}_{\text{资本需求}} + \underbrace{\begin{bmatrix} d_{1,t} \\ d_{2,t} \end{bmatrix}}_{\text{最终需求}}$$

或者

$$Ix_t = Ax_t + C(x_t - x_{t-1}) + d_t$$

但是，将时间下标前移一期，合并同类项，我们可将此方程写 622 成这种形式：

$$(I - A - C)x_{t+1} + Cx_t = d_{t+1} \tag{18.27}$$

当然，微分方程(18.26)和差分方程(18.27)均可由上节提出的方法来解。毫无疑问，将矩阵重新定义并相应改变矩阵维数，就可以将这两个矩阵方程推广至 $n$ 个产业的情况。

上节，我们已经讨论了对时滞和调节机制的考虑是如何引出动态投入产出模型的。把同样的考虑应用于一般均衡市场模型时，它也会以极为相同的方式变成动态模型。但因这种模型的形成与投入产出模型实质上是相似的，所以我们免去形式上的讨论，

读者可以参考练习 18.3－6 和 18.3－7 所介绍的例子。

## 练习 18.3

1　在例 1 中，若最终需求向量变为 $d_t=\begin{bmatrix}\lambda_1\delta^t\\ \lambda_2\delta^t\end{bmatrix}$，那么，特别积分会如何变化？得出答案后，请证明例 1 中的答案仅是在 $\lambda_1=\lambda_2=1$ 时的一个特例。

2　($a$)证明(18.22)可更简洁地写成

$$(\delta I-A)\beta=u$$

($b$)在上面使用的 5 个符号中，哪个是标量？向量？矩阵？

($c$)假设$(\delta I-A)$为非奇异的，以矩阵形式写出 $\beta$ 的解。

3　($a$)证明(18.25)可更简洁地写成

$$(\rho I+I-A)\beta=\lambda$$

($b$)这 5 个符号中，哪些符号分别表示标量、向量或者矩阵？

($c$)假设$(\rho I+I-A)$为非奇异的，以矩阵形式写出 $\beta$ 的解。

4　对于离散时间的生产滞后投入产出模型(18.20)，已知 $A=\begin{bmatrix}\frac{3}{10}&\frac{4}{10}\\ \frac{3}{10}&\frac{2}{10}\end{bmatrix}$ 和 $d_t=\begin{bmatrix}(\frac{12}{10})^t\\ (\frac{12}{10})^t\end{bmatrix}$，求($a$)特别积分；($b$)余函数；($c$)假定初始产出为 $x_{1,0}=\frac{187}{39}$，$x_{2,0}=\frac{72}{13}$，求确定的时间路径。(在所有计算中均使用分数，不要使用小数。)

5　对于离散时间的生产滞后投入产出模型(18.23)，已知 $A=\begin{bmatrix}\frac{3}{10}&\frac{4}{10}\\ \frac{3}{10}&\frac{2}{10}\end{bmatrix}$ 和 $d=\begin{bmatrix}e^{t/10}\\ 2e^{t/10}\end{bmatrix}$，求：($a$)特别积分；($b$)余函数；($c$)假设初始条件为 $x_1(0)=53/6$ 和 $x_2(0)=25/6$，求确定的时间路径。(在所有计算中均使用分数而不使用小数。)

6　在一个 $n$ 种商品的市场中，所有的 $Q_{di}$ 和 $Q_{si}$($i=1,2,\cdots n$)均可视为

$n$ 种价格 $P_1, \cdots, P_n$ 的函数，从而对每一商品的超额需求 $E_i \equiv Q_{di} - Q_{si}$ 也是 623 这些价格的函数。假定函数为线性的，我们可以写出

$$
\begin{aligned}
E_1 &= a_{10} + a_{11}P_1 + a_{12}P_2 + \cdots + a_{1n}P_n \\
E_2 &= a_{20} + a_{21}P_1 + a_{22}P_2 + \cdots + a_{2n}P_n \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
E_n &= a_{n0} + a_{n1}P_1 + a_{n2}P_2 + \cdots + a_{nn}P_n
\end{aligned}
$$

或以矩阵形式

$$E = a + AP$$

($a$)最后这 4 个符号代表什么？标量？向量？还是矩阵？

($b$)将所有价格视为时间的函数，并假定 $dP_i/dt = \alpha_i E_i (i = 1, 2, \cdots, n)$，那么，上面这组方程的经济解释是什么？

($c$)写出表明每个 $dP_i/dt$ 为 $n$ 种价格的线性函数的微分方程。

($d$)证明，若令 $P'$ 表示导数 $dP_i/dt$ 的 $n \times 1$ 列向量，若 $\alpha$ 表示 $n \times n$ 对角矩阵，主对角线上的元素依次为 $\alpha_1, \alpha_2, \cdots, \alpha_n$，其余元素为零，那么，我们可以将上述微分方程组用矩阵符号写成 $P' - \alpha AP = \alpha a$。

7　对于上题中的 $n$ 商品市场，其离散时间形式由一组差分方程 $\Delta P_{i,t} = \alpha_i E_{i,t} (i = 1, 2, \cdots, n)$ 构成，其中 $E_{i,t} = a_{i0} + a_{i1}P_{1,t} + a_{i2}P_{2,t} + \cdots + a_{in}P_{n,t}$。

($a$)写出超额需求方程组，并证明它可以用矩阵符号表示成 $E_t = a + AP_t$。

($b$)证明价格调整方程可以写成 $P_{t+1} - P_t = \alpha E_t$，其中 $\alpha$ 为上题所定义的 $n \times n$ 对角矩阵。

($c$)证明离散时间模型的差分方程组可以以 $P_{t+1} - (I + \alpha A)P_t = \alpha a$ 的形式表示。

---

## 18.4　对通货膨胀—失业模型的进一步讨论

在介绍了多部门投入产出模型的动态方程组以后，我们现在

来介绍一个单部门联立动态方程的经济例子。为此，我们把已在不同题目下遇到过两次的通货膨胀失业模型再一次召回。

### 联立微分方程

在 15.5 节，通货膨胀失业模型是在连续时间框架下通过如下三个方程给出的：

$$p = \alpha - T - \beta U + h\pi \qquad (\alpha, \beta > 0; 0 < h \leqslant 1) \qquad (15.33)$$

$$\frac{d\pi}{dt} = j(p - \pi) \qquad (0 < j \leqslant 1) \qquad (15.34)$$

$$\frac{dU}{dt} = -k(m - p) \qquad (k > 0) \qquad (15.35)$$

那时我们还不具备处理联立动态方程的能力，我们是通过将模型 624 化为一个单变量方程来解决这一问题的，这需要冗繁的代换与消元过程。现在，考虑到在模型中共存 $\pi$ 和 $U$ 的两个已知的变化模式，我们将其作为一个由两个联立微分方程构成的模型。

把(15.33)代入另两个方程并把导数 $d\pi/dt \equiv \pi'(t)$ 和 $dU/dt \equiv U'(t)$ 更简洁地写成 $\pi'$ 和 $U'$ 后，模型的形式为

$$\begin{aligned} \pi' + j(1-h)\pi + j\beta U &= j(\alpha - T) \\ U' - kh\pi + k\beta U &= k(\alpha - T - m) \end{aligned} \qquad (18.28)$$

或者用矩阵符号

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{J} \begin{bmatrix} \pi' \\ U' \end{bmatrix} + \underbrace{\begin{bmatrix} j(1-h) & j\beta \\ -kh & k\beta \end{bmatrix}}_{M} \begin{bmatrix} \pi \\ U \end{bmatrix} = \begin{bmatrix} j(\alpha - T) \\ k(\alpha - T - m) \end{bmatrix} \qquad (18.28')$$

由此方程组，可立即求得 $\pi$ 和 $U$ 的时间路径。而且，如果需要的

话，我们还可以通过(15.33)导出 $p$ 的路径。

## 解的路径

为求得特别积分，我们可令(18.28′)中的$\pi' = U' = 0$(使$\pi$和$U$对时间稳定)并解出$\pi$和 $U$。在前面(18.14)的讨论中，这样的解是通过矩阵求逆的方法求出的，当然克莱姆法则必然是要用到的。运用这两种方法中的哪一种，我们都能求得

$$\bar{\pi} = m \quad 和 \quad \bar{U} = \frac{1}{\beta}[\alpha - T - (1-h)m] \tag{18.29}$$

$\bar{\pi} = m$(预期的均衡通货膨胀率等于货币扩张率)这个结果与15.5节所得到的结果是一致的。至于失业率 $U$，我们在本节并不准备求出其均衡水平。但如果我们作这种尝试的话(在练习15.5-2给出的 $U$ 的微分方程的基础上)，答案与(18.29)中的解 $\bar{U}$没什么差别。

现在转到余函数，这些余函数是以试探解 $me^{rt}$和 $ne^{rt}$为基础的，由下列简化矩阵方程我们可以确定 $m$、$n$ 和$r$：

$$(rJ + M)\begin{bmatrix} m \\ n \end{bmatrix} = 0 \quad [源于(18.15)]$$

在这里，它取如下形式

$$\begin{bmatrix} r + j(1-h) & j\beta \\ -kh & r + k\beta \end{bmatrix}\begin{bmatrix} m \\ n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{18.30}$$

625 为避免 $m$ 和 $n$ 的试探解偏离此齐次方程组，系数矩阵的行列式必定为零。即我们要求

$$|rJ + M| = r^2 + [k\beta + j(1-h)]r + k\beta j = 0 \tag{18.31}$$

此二次方程是特征方程 $r^2 + a_1 r + a_2 = 0$ 的一个特殊形式，其系数为

$$a_1 = k\beta + j(1 - h) \quad 和 \quad a_2 = k\beta j$$

而且正如我们预期的那样，这两个系数恰好是(15.37″)中的 $a_1$ 和 $a_2$ 值。而(15.37″)是我们现在模型的以$\pi$ 为变量的单一方程形式。因此，我们以前关于特征根三种情况的分析应用到这里应当是同样成立的。在其它的结论中，我们还记得，无论根恰好是实根还是复根，在现在的模型中根的实的部分将总是负的。因此解的路径总是收敛的。

**例 1** 已知参数值

$$a - T = \frac{1}{6} \quad \beta = 3 \quad h = 1 \quad j = \frac{3}{4} \quad k = \frac{1}{2}$$

求$\pi$ 和 $U$ 的时间路径。因为这些参数值与 15.5 节中例 1 的参数值相同，所以现在的分析结果可以用于验证 15.5 节的结果。

首先，容易确定特别积分为

$$\bar{\pi} = m \quad 和 \quad \bar{U} = \frac{1}{3}\left(\frac{1}{6}\right) = \frac{1}{18} \quad [由(18.29)] \tag{18.32}$$

特征方程为

$$r^2 + \frac{3}{2}r + \frac{9}{8} = 0 \quad [由(18.31)]$$

两个根为复根

$$r_1, r_2 = \frac{1}{2}\left(-\frac{3}{2} \pm \sqrt{\frac{9}{4} - \frac{9}{2}}\right) = -\frac{3}{4} \pm \frac{3}{4}i \tag{18.33}$$

$$\left(其中\ h = -\frac{3}{4}\ 和\ v = \frac{3}{4}\right)$$

将两个根与参数值一道代入(18. 30),分别得到矩阵方程

$$\begin{bmatrix} -\frac{3}{4}(1-i) & \frac{9}{4} \\ -\frac{1}{2} & \frac{3}{4}(1+i) \end{bmatrix} \begin{bmatrix} m_1 \\ n_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \left[\text{由 } r_1 = -\frac{3}{4} + \frac{3}{4}i\right] \tag{18.34}$$

$$\begin{bmatrix} -\frac{3}{4}(1+i) & \frac{9}{4} \\ -\frac{1}{2} & \frac{3}{4}(1-i) \end{bmatrix} \begin{bmatrix} m_1 \\ n_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \left[\text{由 } r_2 = -\frac{3}{4} - \frac{3}{4}i\right] \tag{18.34'}$$

626 因为通过(18.31),$r_1$ 和 $r_2$ 被设计成使系数矩阵为奇异矩阵,所以,上述两个矩阵方程中的每一个方程实际上只包含一个独立方程,它可以在任意常数 $m_i$ 和 $n_i$ 之间确定一种比例关系。具体地,我们有

$$\frac{1}{3}(1-i)m_1 = n_1 \quad \text{和} \quad \frac{1}{3}(1+i)m_2 = n_2$$

相应地,余函数可以表示成

$$\begin{bmatrix} \pi_c \\ U_c \end{bmatrix} = \begin{bmatrix} m_1 e^{r_1 t} + m_2 e^{r_2 t} \\ n_1 e^{r_1 t} + n_2 e^{r_2 t} \end{bmatrix}$$

$$= e^{ht} \begin{bmatrix} m_1 e^{vit} + m_2 e^{-vit} \\ n_1 e^{vit} + n_2 e^{-vit} \end{bmatrix} \qquad [\text{由}(15.11)]$$

$$= e^{ht} \begin{bmatrix} (m_1 + m_2)\cos vt + (m_1 - m_2) i \sin vt \\ (n_1 + n_2)\cos vt + (n_1 - n_2) i \sin vt \end{bmatrix}$$

$$[\text{由}(15.24)]$$

为使记法简单,如果我们定义新的任意常数

$$A_5 \equiv m_1 + m_2 \quad \text{和} \quad A_6 \equiv (m_1 - m_2)i$$

则由此可得[①]

$$n_1 + n_2 = \frac{1}{3}(A_5 - A_6) \quad (n_1 - n_2)i = \frac{1}{3}(A_5 + A_6)$$

运用这些结果,并将(18.33)的 $h$ 和 $v$ 值纳入余函数,从而得到

$$\begin{bmatrix} \pi_c \\ U_c \end{bmatrix} = e^{-3t/4}\begin{bmatrix} A_5\cos\frac{3}{4}t + A_6\sin\frac{3}{4}t \\ \frac{1}{3}(A_5 - A_6)\cos\frac{3}{4}t + \frac{1}{3}(A_5 + A_6)\sin\frac{3}{4}t \end{bmatrix} \tag{18.35}$$

最后,将特别积分(18.32)与上述余函数合并,我们可以得到$\pi$和 $U$ 的解路径。正如所预期的那样,这些路径与 15.5 节中的路径(15.43)和(15.45)完全相同。

## 联立差分方程

627

离散时间条件下通货膨胀—失业模型的处理与上一节在连续时间情况下的讨论是大致相似的。因此,我们这里仅给出其要点。

如17.3节给出的那样,要研究的模型由三个方程组成,其中

---

① 此结论可由下式看出:

$$\begin{aligned} n_1 + n_2 &= \frac{1}{3}(1-i)m_1 + \frac{1}{3}(1+i)m_2 \\ &= \frac{1}{3}[(m_1 + m_2) - (m_1 - m_2)i] \\ &= \frac{1}{3}(A_5 - A_6) \\ (n_1 - n_2)i &= \left[\frac{1}{3}(1-i)m_1 - \frac{1}{3}(1+i)m_2\right]i \\ &= \frac{1}{3}[(m_1 - m_2) - (m_1 + m_2)i]i \\ &= \frac{1}{3}(A_6 + A_5) \quad [i^2 \equiv -1] \end{aligned}$$

的两个方程分别描述$\pi$和 $U$ 的变化模式：

$$p_t = \alpha - T - \beta U_t + h\pi_t \tag{17.18}$$

$$\pi_{t+1} - \pi_t = j(p_t - \pi_t) \tag{17.19}$$

$$U_{t+1} - U_t = -k(m - p_{t+1}) \tag{17.20}$$

消去 $p$，合并同类项，我们可以将模型重写成差分方程组：

$$\underbrace{\begin{bmatrix} 1 & 0 \\ -kh & 1+\beta k \end{bmatrix}}_{J} \begin{bmatrix} \pi_{t+1} \\ U_{t+1} \end{bmatrix} + \underbrace{\begin{bmatrix} -(1-j+jh) & j\beta \\ 0 & -1 \end{bmatrix}}_{K} \begin{bmatrix} \pi_t \\ U_t \end{bmatrix} = \begin{bmatrix} j(\alpha - T) \\ k(\alpha - T - m) \end{bmatrix} \tag{18.36}$$

## 解的路径

如果存在静态均衡，(18.36)的特别积分可以表示成$\bar{\pi} = \pi_t = \pi_{t+1}$和 $\bar{U} = U_t = U_{t+1}$。将$\bar{\pi}$和 $\bar{U}$代入(18.36)，并通过矩阵求逆或克莱姆法则解方程组，得到

$$\bar{\pi} = m \quad 和 \quad \bar{U} = \frac{1}{\beta}[\alpha - T - (1-h)m] \tag{18.37}$$

$\bar{U}$值与我们在 17.3 节所求到的值是相同的。尽管我们在 17.3 节中没有求出$\bar{\pi}$值，但练习 17.3－2 中的信息表明$\bar{\pi} = m$，这与(18.37)是一致的。读者或许注意到，结果(18.37)与在连续时间框架(18.29)中得到的瞬时均衡值也是一致的。

求以试探解 $mb^t$ 和 $nb^t$ 为基础的余函数涉及简化矩阵方程

$$(bJ + K)\begin{bmatrix} m \\ n \end{bmatrix} = 0$$

或考虑到(18.36)

$$\begin{bmatrix} b-(1-j+jh) & j\beta \\ -bkh & b(1+\beta k)-1 \end{bmatrix}\begin{bmatrix} m \\ n \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (18.38)$$

为避免试探解偏离此齐次方程组,我们要求

$$|bJ+K|=(1+\beta k)b^2-[1+hj+(1-j)(1+\beta k)]b$$
$$+(1-j+jh)=0 \quad (18.39)$$

此二次方程的正规化形式是特征方程 $b^2+a_1b+a_2=0$,系数 $a_1$ 和 $a_2$ 628 与 17.3 节中(17.24)及(17.33)的系数相同。因而,17.3 节对特征根三种情况的分析应同样适用于这里。

对于每个根 $b_i$,(18.38)为我们提供了任意常数 $m_i$ 与 $n_i$ 之间的一个特定的比例关系,而且它们可以使我们将 $U$ 的余函数中的任意常数与$\pi$的余函数中的任意常数联系起来。进而,把余函数和特别积分合并,便可以得到$\pi$和 $U$ 的时间路径。

## 练习 18.4

---

1 运用克莱姆法则验证(18.29)。

2 验证:在方程组(18.34)中,无论我们使用一阶方程还是二阶方程,$m_1$ 和 $m_2$ 之间都会存在同样的比例关系。

3 求$\pi$和 $U$ 的时间路径(通解),已知

(a) $p=\frac{1}{4}-2U+\pi$

$\pi'=\frac{1}{2}(p-\pi)$

$U'=-(m-p)$

(b) $p=\frac{1}{6}-2U+\frac{1}{3}\pi$

$\pi'=\frac{1}{4}(p-\pi)$

$U'=-\frac{1}{2}(m-p)$

4 求$\pi$和 $U$ 的时间路径(通解),已知:

(a) $p_t=\frac{1}{2}-3U_t+\frac{1}{2}\pi_t$ (b) $p_t=\frac{1}{4}-4U_t+\pi_t$

$\pi_{t+1}-\pi_t=\frac{1}{4}(p_t-\pi_t)$ $\pi_{t+1}-\pi_t=\frac{1}{4}(p_t-\pi_t)$

$U_{t+1}-U_t=-(m-p_{t+1})$ $U_{t+1}-U_t=-(m-p_{t+1})$

---

## 18.5 双变量相位图

上节我们研究了如何获得线性动态方程组的定量解。在本节，我们将讨论非线性微分方程组的定性图解(相位图)分析。更确切地，我们将把注意力集中于双变量一阶微分方程组，其一般形式为

$$x'(t)=f(x,y)$$
$$y'(t)=g(x,y)$$

注意，时间导数 $x'(t)$和 $y'(t)$仅取决于 $x$ 和 $y$，变量 $t$ 并不作为独立变数进入 $f$ 和 $g$ 函数。这一特征使方程组成为一个自控系统。它是应用相位图分析方法的前提条件。

629 同 14.6 节介绍的单变量相位图一样，双变量相位图的局限性在于，它只能回答定性的问题，即那些关于瞬时均衡的位置及动态稳定性的问题。但是，依旧同单变量相位图一样，它也具有一个补偿性的优势，即它可以像处理线性问题一样方便地处理非线性方程组，而且，能够像用具体函数形式一样容易地用一般函数形式表达问题。

### 相空间

当构建一个(自控)微分方程 $dy/dt=f(y)$的单变量相位图时，我们只需在两维空间中的两轴上对 $y$ 绘出 $dy/dt$ 即可。但现在变量数

加倍了，我们如何安排，才能满足对更多坐标轴的明显需要呢？幸运的是，二维空间已经完全能够满足这种需要。

为了解为什么二维空间是可行的，需要认识到构建相位图的最重要的任务，是确定变量随时间而运动的方向。正如图 14.3 的箭头所表示的那样，正是运动方向这一信息，使我们得以推断出最终的定性结论。因为绘出所说的箭头，只需要两个条件：(1)一条分界线，称之为"$dy/dt=0$"线，它为任何预期的均衡提供一个发生的地点；更重要的是，它把相空间分成两个区域，一个为 $dy/dt>0$ 的区域，另一个为 $dy/dt<0$ 的区域；(2)一条实线。在此线上，由 $dy/dt$ 的任意非零值所反映的 $y$ 的递增和递减可以标示出来。在图 14.3 中，上面第一项所说的分界线是横轴，但实际上横轴也是上面第二点所说的实线。这意味着如果我们关注的是在 $dy/dt>0$ 的区域和 $dy/dt<0$ 的区域间作出区分，比如给前者加上加号，给后者加上减号，那么，放弃纵轴 $dy/dt$ 也不会产生什么负面影响。一个轴的可省略性使得在二维空间中置放双变量相位图成为可行。我们现在需要两条实线而非一条实线，而二维图形中的标准的 $x$ 轴和 $y$ 轴可以自动地满足这一点。我们还需要两条分界线，一条是 $dx/dt=0$ 线，另一条是 $dy/dt=0$ 线。而这两条线在二维相空间中均是可以绘出的。而一旦绘出这些分界线，确定在这些线的哪一边应标上加号，哪一边应标上减号，也就不再困难了。

## 分界曲线

已知如下自控微分方程组

$$
\begin{aligned}
x' &= f(x,y) \\
y' &= g(x,y)
\end{aligned}
\tag{18.40}
$$

630 其中 $x'$和 $y'$为时间导数 $x'(t)$和 $y'(t)$的简写形式,两条分界线(以 $x'=0$ 和 $y'=0$ 表示)表示如下两个方程的图形:

$$f(x,y)=0 \quad [x'=0\text{ 曲线}] \tag{18.41}$$

$$g(x,y)=0 \quad [y'=0\text{ 曲线}] \tag{18.42}$$

如果 $f$ 函数的具体形式为已知,那么,由(18.41)可对 $x$ 解出 $y$,而且绘在 $xy$ 平面上的解可以作为 $x'=0$ 曲线。即便不能,我们还可以借助于隐函数法则,确定 $x'=0$ 曲线的斜率为

$$\left.\frac{dy}{dx}\right|_{x'=0}=-\frac{\partial f/\partial x}{\partial f/\partial y}=-\frac{f_x}{f_y} \quad (f_y\neq 0) \tag{18.43}$$

只要知道偏导数 $f_x$ 和 $f_y(\neq 0)$的符号,就可以由(18.43)得到$x'=0$ 曲线斜率的定性线索。同理,$y'=0$ 曲线的斜率可由导数

$$\left.\frac{dy}{dx}\right|_{y'=0}=-\frac{g_x}{g_y} \quad (g_y\neq 0) \tag{18.44}$$

推断出来。

为介绍得更具体,我们假设

$$f_x<0 \quad f_y>0 \quad g_x>0 \quad \text{和} \quad g_y<0 \tag{18.45}$$

则 $x'=0$ 和 $y'=0$ 曲线的斜率均为正。若进一步假定

$$-\frac{f_x}{f_y}>-\frac{g_x}{g_y} \quad [x'=0\text{ 曲线陡于 }y'=0\text{ 曲线}]$$

则会出现如图 18.1 所示的情形。注意现在分界线可能是曲线。还要注意,现在不再要求它们与轴重合。

两条分界线在点 $E$ 相交,将相空间分成四个标记为 I 到 IV 的四个区域。点 $E$(此处 $x$ 和 $y$ 均为稳定的,$x'=y'=0$)表示方程组的瞬时均衡。但在其它任何点,无论 $x$ 还是 $y$,都会随时间而变化,变化的方向由时间导数 $x'$和 $y'$在该点的符号来确定。在本例中,我们恰好

在 $x'=0$ 曲线的左边(右边)有 $x'>0(x'<0)$,因此该曲线左边(右边)为加号(减号)。这些符号以下述事实为基础

$$\frac{\partial x'}{\partial x} = f_x < 0 \quad [\text{由(18.40)和(18.45)}] \tag{18.46}$$

它意味着当我们在相位空间中持续由西向东移动时(当 $x$ 增加时),$x'$ 稳定减少,从而 $x'$ 的符号必定依次经过三个阶段:+,0,-。类似地,导数

$$\frac{\partial y'}{\partial y} = g_y < 0 \quad [\text{由(18.40)和(18.45)}] \tag{18.47}$$

意味着当我们持续由南向北移动时(当 $y$ 增加时),$y'$ 稳定减少,从而 631
$y'$ 的符号必定依次经过三个阶段:+,0,-。因而我们在图 18.1 $y'=0$ 曲线的下面加上加号,在其上面加上减号。

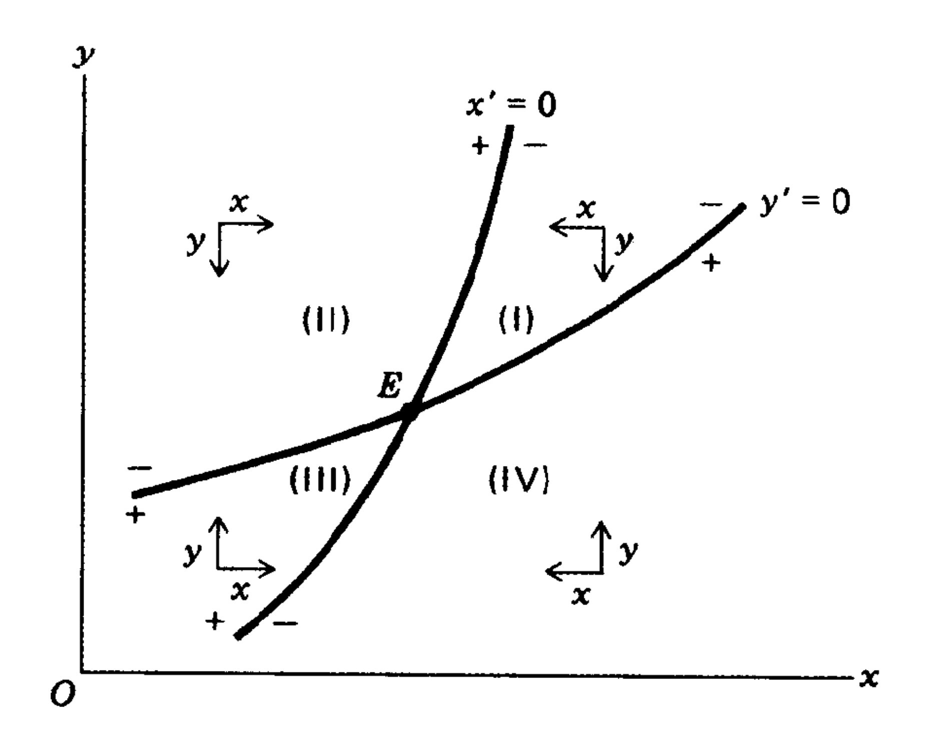



图 18.1

在这些加减号的基础上,我们可以绘出一组带有方向的箭头,以指示 $x$ 和 $y$ 的瞬时运动方向。在区域 I 中的任意点,$x'$ 与 $y'$ 均为负,

因而 $x'$ 与 $y'$ 必随时间而递减，$x$ 会向西运动，$y$ 则向南运动。如图区域 I 中的两个箭头所示，给定区域 I 中的起始点，瞬时运动的一般方向必然是指向西南。区域 III 的情况恰好与区域 I 相反，$x'$、$y'$ 均为正，$x$、$y$ 变量必随时间而增加。在区域 II 中，$x'$、$y'$ 的符号不同，$x'$ 为正，$y'$ 为负，$x$ 应向东移动，$y$ 应向南移动。而区域 IV 则表现出与区域 II 恰好相反的趋势。

## 流　线

为更好地理解方向箭头的含义，我们可在相位图中绘出一系列的流线。这些流线也称作相轨道(*phase trajectories*)，简称轨道，或者相路径(*phase paths*)，它们可以表示出系统从任意可接受的起始点的动态运动。我们在图 18.2 中绘出了一些流线，其中的 $x'=0$ 曲线与 $y'=0$ 曲线与(18.1)中的曲线相同。因为相空间中的每一点必定位于某一流线上，所以应存在无数条流线，所有的流线均与每一区域中 $xy$ 箭头所规定的方向要求相一致。但对于描述相位图的一般定性特征而言，632 少数代表性的流线就已足够了。

对图 18.2 中的流线，应注意以下几个特征。首先，所有流线均指向 $E$ 点。这使得 $E$ 成为一个稳定(这里是整体稳定)瞬时均衡点。后面，我们还会遇到其它类型的流线图形。其次，尽管某些流线不会位于某一单个区域之外(如通过点 $A$ 的流线)，但另外一些流线却可能从一个区域穿过另一个区域(比如通过点 $B$ 和 $C$ 的两条流线)。第三，其中一条跨区域的流线必定或者具有无穷大的斜率(跨越 $x'=0$ 曲线)，或者具有零斜率(跨越 $y'=0$ 线)。这是因为沿着 $x'=0(y'=0)$ 线，$x(y)$ 对时间稳定，因而当穿过 $x'=0(y'=0)$ 线时，流线必定不会有任何水平(垂直)移动。为保证这些斜率要求始终能够得到满足，只

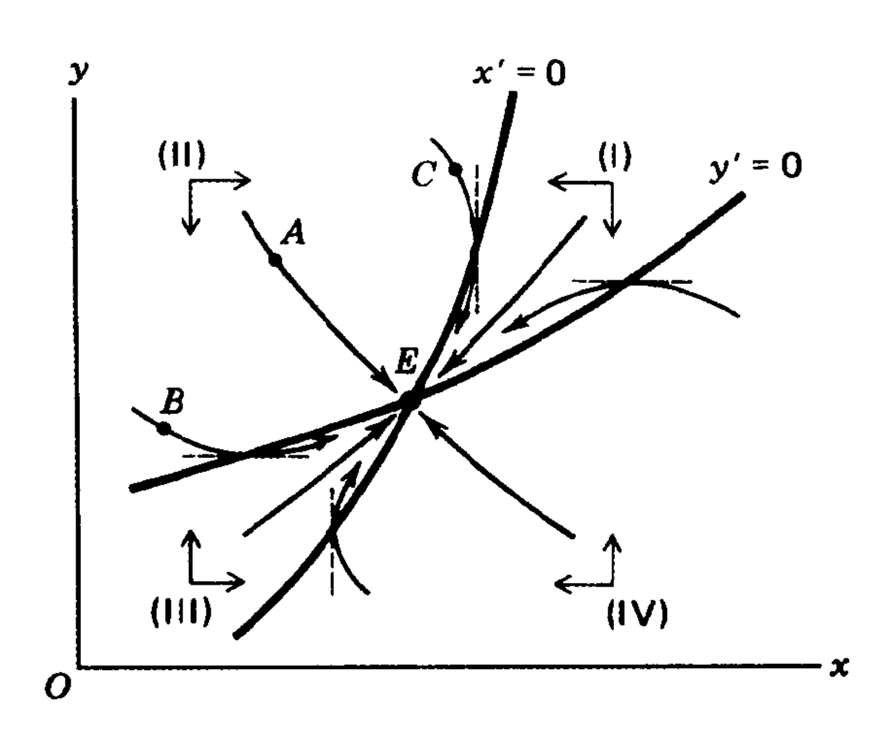



图 18.2

要一把分界线放在合适的位置上，便在穿越 $x'=0$ 曲线处加上一条短的垂直虚线，在穿越 $y'=0$ 曲线处加上一条短的水平虚线，以作为绘制流线的指引。① 第四，尽管流线确实明确地指出了 $x$、$y$ 随时间运动的方向，但它们不能提供任何关于运动速度与加速度的具体信息，因为相位图中并不存在 $t$(时间)轴。当然，正是由于这一原因，流线有了另一个名称**相路径**，以与**时间路径**相对照。关于速度，我们能观察到 633 的唯一信息本质上是定性的：当我们沿着流线越来越趋近于 $x'=0$($y'=0$)线时，在水平(垂直)方向趋近的速度必定逐渐递减。这是因为当我们向 $x'$($y'$)取零值的分界线移动时，导数 $x'\equiv dx/dt$($y'\equiv dy/dt$)的绝对值会稳定地减少。

① 为帮助记忆，应注意与 $x'=0$ 曲线交叉的虚线应垂直于 $x$ 轴。类似地，与 $y'=0$ 交叉的虚线应垂直于 $y$ 轴。

## 均衡的类型

根据围绕特定瞬时均衡的流线的形状，我们可以把均衡分成四种类型：(1)结点；(2)鞍点；(3)焦点；(4)涡旋。

结点均衡是一种使得所有与之相联系的流线或者是非循环地流向均衡点(稳定结点)，或者是非循环地背离均衡点(非稳定结点)的均衡。在图 18.2 中我们已遇到了稳定结点均衡，而非稳定结点均衡则如图 18.3$a$ 所示。注意，在图 18.3$a$ 这个特定的情况下，流线恰好不跨越不同的区域，而且 $x'=0$，$y'=0$ 恰好为直线。事实上，它们本身也恰好为流线。

鞍点均衡是一种具有双重特性的均衡：它在某些方向是稳定的，在另一些方向是不稳定的。更确切地说，参照图 18.3$b$，一个鞍点均衡恰好具有一对流线(称作鞍点的稳定枝)始终直接流向均衡点，恰好有另一对流线(非稳定枝)始终直接背离均衡点。所有其它最初指向鞍点的轨道迟早都要背离它。当然，正是这种双重特性使其得到了"鞍点"这一名称。因为稳定性只有从稳定枝才能观察到，它并不具备必然性，所以鞍点均衡一般被视为非稳定均衡。

第三种类型的均衡是焦点均衡。其特征是具有涡旋轨道，这些涡旋轨道或者循环性地流向焦点(稳定焦点)，或者是循环性地背离焦点(非稳定焦点)。图 18.3$c$ 描绘了一个稳定焦点，只明确地绘出一条流线以避免凌乱。是何因素导致涡旋运动呢？答案在于 $x'=0$，$y'=0$ 曲线定位的方式。在图 18.3$c$ 中，两条分界线是以这样一种方式倾斜的：他们依次封锁流线按特定的 $xy$ 箭头所确定的方向流动，结果流线被迫从一个区域穿过另一个区域，作螺旋形的运动。我们是能得到一个稳定的焦点平衡(如这里的情况)，还是一个非稳定的焦点平衡，取

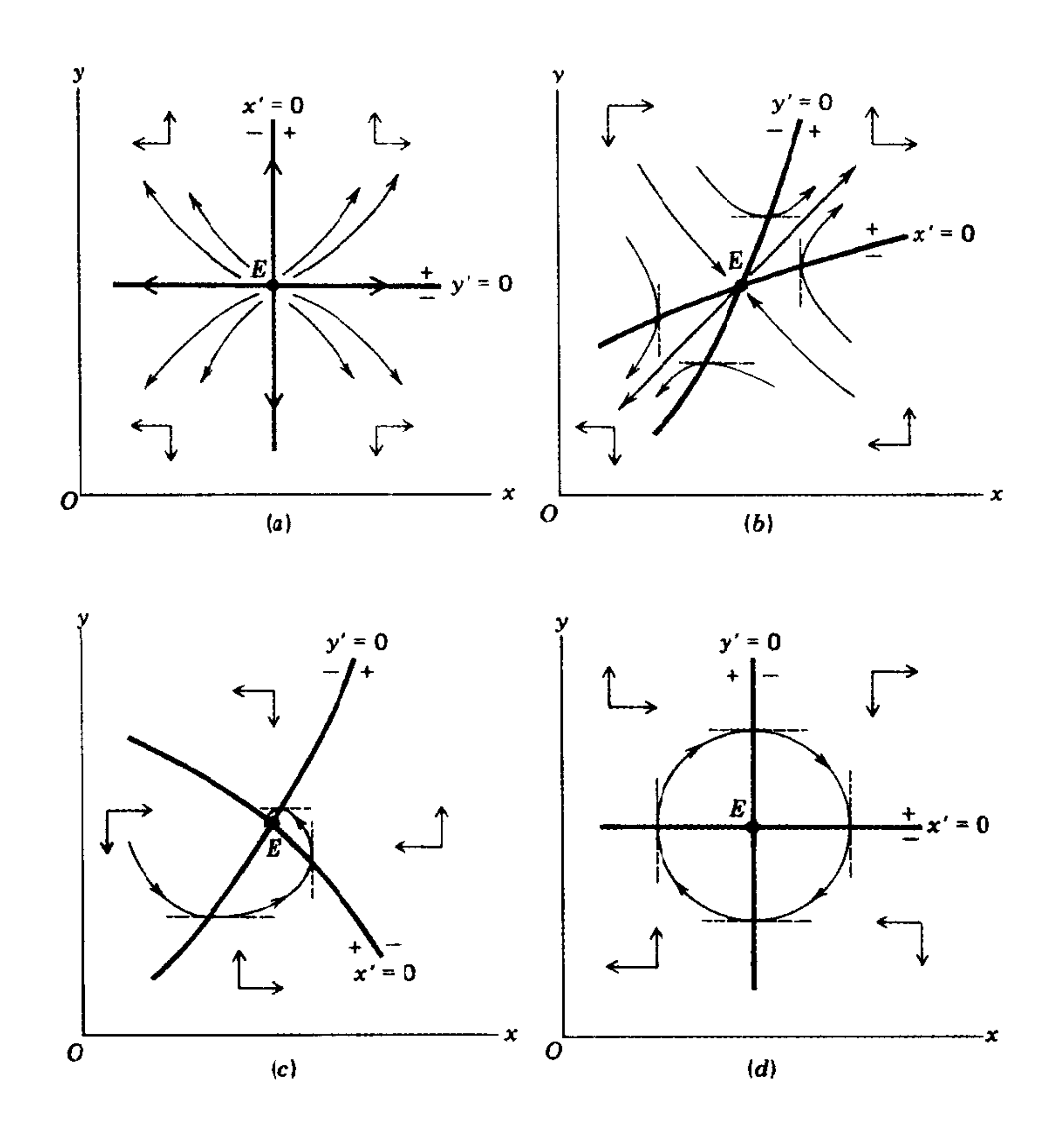



图 18.3

决于两条分界线的相对位置。但无论在哪种情况下，流线在交叉点的斜率必定或者是无穷大（交于 $x'=0$），或者是零（交于 $y'=0$）。

最后是涡旋（中心）均衡。这种均衡同样具有涡旋流线，但这些流线现在形成一组环形（同心圆或椭圆），围绕着均衡进行永恒的运动。在图 18.3$d$ 中给出了这种均衡的一个例子，但在图中我们依然只绘出了一条流线。由于在 $E$ 点以外的任何初始位置均不可能达到这种均衡，所以涡旋均衡自然属于非稳定均衡。

634 图 18.3 所介绍的情况只有唯一均衡。但若存在足够的非线性，两条分界线会交叉多次，从而产生多重均衡。在这种情况下，上面所述的不同类型的瞬时均衡的组合可能会出现在一个相位图中。尽管这样可能会出现四个以上的区域，但相位图分析所依赖的主要原理依然是基本相同的。

## 635 奥比斯特的通货膨胀与货币规则

为给出一个两变量相位图的经济例子，我们介绍一个奥比斯特教授提出的模型。[①] 此模型的目的是证明在通货膨胀调节机制发挥作用的情况下，传统的反周期货币政策规则是无效的，因而需要新的货币政策规则。此模型与我们前面对通货膨胀的讨论相反，它不去研究给定货币扩张率的影响，而是考察两种不同货币规则的功效。每种货币规则都描述在面临不同通胀条件下所采取的一组不同的货币行动。

此模型的一个关键假设是通货膨胀调节机制

$$\frac{dp}{dt} = h\left(\frac{M_s - M_d}{M_s}\right) = h\left(1 - \frac{M_d}{M_s}\right) \quad (h > 0) \qquad (18.48)$$

此式表明，超额货币供给的效果在于提高通货膨胀率 $p$，而非价格水平 $P$。因而，货币市场出清不意味着价格的稳定，而是意味着一个稳定的通货膨胀率。为便于分析，(18.48)中的第二个等式可帮助我们将注意力由超额货币供给转移到货币的需求—供给比率 $M_d/M_s$ 上。我们后面将以 $\mu$ 表示 $M_d/M_s$。假设 $M_d$ 直接与名义国民产品 $PQ$ 成比例，则可写成

① 诺曼·奥比斯特："具有通货膨胀调节机制的稳定政策"，载《经济学季刊》，1978 年 5 月，第 355—359 页。奥比斯特的文章中没有相位图，但可由其模型很容易地构建相位图。

$$\mu \equiv \frac{M_d}{M_s} = \frac{aPQ}{M_s} \quad (a > 0)$$

几个相关变量的增长率则可以由

$$\frac{d\mu/dt}{\mu} = \frac{da/dt}{a} + \frac{dP/dt}{P} + \frac{dQ/dt}{Q} - \frac{dM_s/dt}{M_s}$$

［由(10.24)和(10.25)］

$$\equiv p + q - m \quad [a = \text{常数}] \tag{18.49}$$

联系起来，其中小写字母 $p$，$q$ 和 $m$ 分别表示通货膨胀率、外生的实际国民产品增长率和货币扩张率。

如果这次我们把 $m$ 视为外生的，那么方程(18.48)和(18.49)，两个微分方程的组合，可共同确定 $p$ 和 $\mu$ 的时间路径。使用符号 $p'$ 和 $\mu'$ 表示时间导数 $p'(t)$ 和 $\mu'(t)$，我们可将方程组更简洁地表示成

$$\begin{aligned} p' &= h(1-\mu) \\ \mu' &= (p+q-m)\mu \end{aligned} \tag{18.50}$$

给定 $h$ 为正，当且仅当 $1-\mu=0$ 时，我们可以有 $p'=0$。类似地，因为 $\mu$ 总为正，所以当且仅当 $p+q-m=0$ 时，$\mu'=0$。因此 $p'=0$ 和 $\mu'=0$ 636 分界线与方程

$$\mu = 1 \qquad [p' = 0 \text{ 曲线}] \tag{18.51}$$

$$p = m - q \qquad [\mu' = 0 \text{ 曲线}] \tag{18.52}$$

相联系。如图 18.4$a$ 所示，分界线被分别绘成水平线和垂直线，在 $E$ 点产生唯一的均衡。均衡值 $\overline{\mu} = 1$ 意味着均衡时的 $M_d$ 与 $M_s$ 相等，从而使货币市场出清。均衡的通货膨胀率表示为正这一事实反映了一个隐含的假设：$m > q$。

因为 $p'=0$ 曲线对应于我们前面讨论中的 $x'=0$ 曲线，所以它应

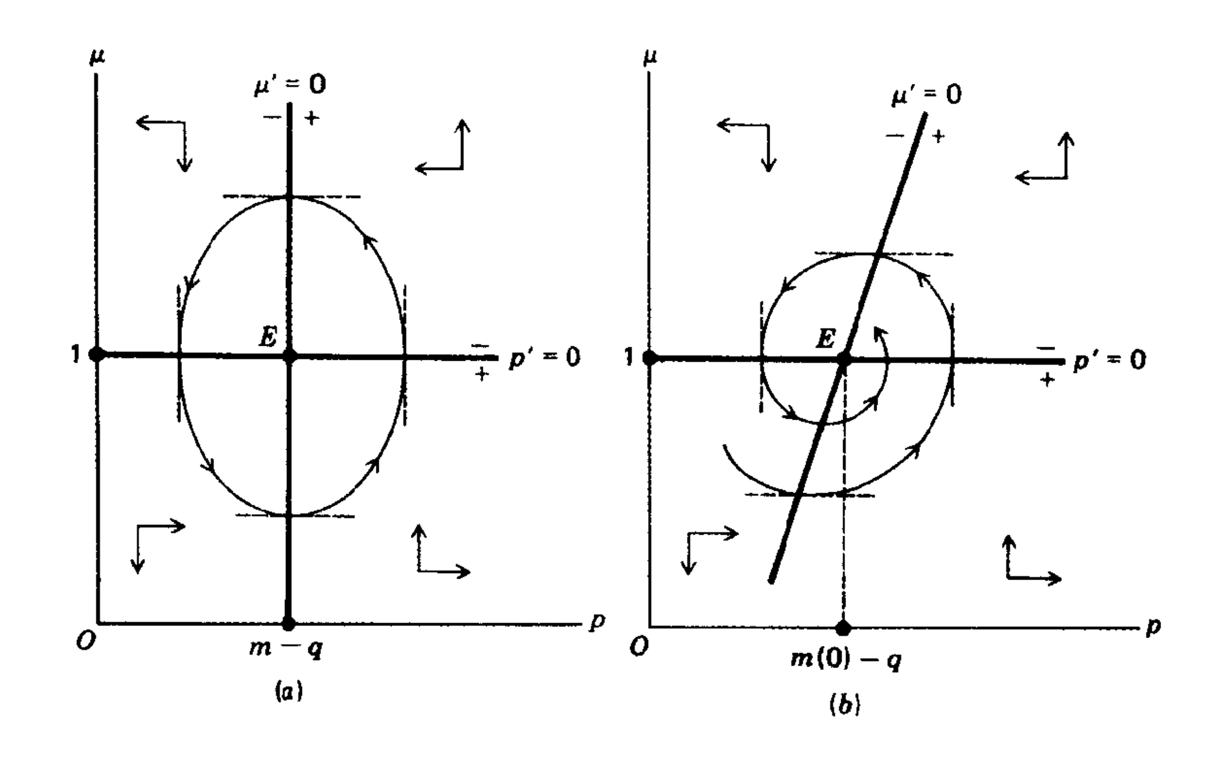



图 18.4

具有垂直的虚线。而另一分界线则应具有水平的虚线。由(18.50)，可以求得

$$\frac{\partial p'}{\partial \mu} = -h < 0 \quad 和 \quad \frac{\partial \mu'}{\partial p} = \mu > 0 \tag{18.53}$$

它表明这样一种含义：穿过 $p'=0$ 曲线向北移动，$p'$ 的符号依次通过(+,0,−)区域；穿过 $\mu'=0$ 曲线向东移动，$\mu'$ 的符号依次通过(−,0,+)区域。因此，我们获得四组如图所示的方向箭头，它们产生围绕均衡点 $E$ 逆时针旋转的流线(图中仅绘出一条)。这使得 $E$ 成为涡旋均衡。除非该经济恰好始于 $E$ 点，否则不可能达到均衡。它们会处于永不停息的波动状态。

637 但是，上面的结论是外生货币扩张率的结果。如果我们通过反通货膨胀货币规则使 $m$ 内生化，会出现何结果呢？"传统"货币规则要求货币扩张率与通货膨胀率反向运动：

$$m = m(p) \quad m'(p) < 0 \quad [传统的货币规则] \tag{18.54}$$

此规则使(18.50)中的第二个方程修正为

$$\mu' = [p + q - m(p)]\mu \tag{18.55}$$

并把(18.52)变为

$$p = m(p) - q \quad [传统货币规则下的 \mu' = 0 曲线] \tag{18.56}$$

已知 $m(p)$是单调的,仅存在一个 $p$ 值(比如 $p_1$)可以满足此方程。因此新的 $\mu'=0$ 曲线必定会是一条垂线,尽管具有不同的水平的交点 $p_1 = m(p_1) - q$。进而,由(18.55),可以求得

$$\frac{\partial \mu'}{\partial p} = [1 - m'(p)]\mu > 0 \quad [由(18.54)]$$

它与(18.53)中的导数在定性意义上并无差别。由此可知,其方向箭头必然也与(18.4$a$)中的方向箭头相同。总之,我们以得到一个同以前一样的涡旋均衡而告终。

奥比斯特所提出的另一货币规则是,使 $m$ 与通货膨胀率的变化率(而不是其水平)相协调:

$$m = m(p') \quad m'(p') < 0 \quad [另一种货币规则] \tag{18.57}$$

在此规则下,(18.55)和(18.56)分别变成

$$\mu' = [p + q - m(p')]\mu \tag{18.58}$$

$$p = m(p') - q \quad [另一种货币规则下的 \mu' = 0 曲线] \tag{15.59}$$

这次 $\mu'=0$ 线变成一条向上倾斜的曲线。根据链式法则将(18.59)对 $\mu$ 微分,得到

$$\frac{dp}{d\mu} = m'(p')\frac{dp'}{d\mu} = m'(p')(-h) > 0 \quad [由(18.50)]$$

根据反函数法则,$d\mu/dp$($\mu'=0$ 线的斜率)也为正。这种新的情况被描绘在图 18.4$b$ 中,其中为简便起见,$\mu'=0$ 曲线被绘成一条直线,它

具有一个任意确定的斜率。[1] 尽管斜率是变化的，由(18.53)知，导数

$$\frac{\partial \mu'}{\partial p} = \mu > 0 \quad [源于(18.58)]$$

是不变的，从而 $\mu$ 的符号应与图 18.4$a$ 中原来的方向相同。流线(图中仅绘出一条)将朝着在 $\bar{\mu}=1$ 和 $\bar{p}=m(0)-q$ 的均衡点向内旋转，其中 $m(0)$ 表示在 $p'=0$ 计值的 $m(p')$。因此，这种货币规则看起来能够将涡旋均衡转化为稳定的焦点均衡，从而有可能逐渐地消除通货 638 膨胀率的永久波动。事实上，如果具有足够平坦的 $\mu'=0$ 曲线，甚至有可能将涡旋均衡转化为稳定的结点均衡。

## 练习 18.5

---

1 证明：如果模型仅由一个二阶微分方程 $y''(t)=f(y',y)$ 构成，而不是由两个一阶方程构成，那么，仍可运用两变量相位图分析。

2 图 18.1 中 $x'=0$ 和 $y'=0$ 曲线两边所附的加号和减号是分别建立在偏导数 $\partial x'/\partial x$ 和 $\partial y'/\partial y$ 基础上的。可以由导数 $\partial x'/\partial y$ 和 $\partial y'/\partial x$ 得出同样的结论吗？

3 运用图 18.2 验证在穿过 $x'=0(y'=0)$ 曲线时，如果流线不具有无穷大(零)的斜率，就必然会违背由 $xy$ 箭头所施加的方向限制。

4 作为微分方程组(18.40)的一个特例，假设

($a$) $f_x=0$ $\quad f_y>0 \quad g_x>0$ 和 $\quad g_y=0$

($b$) $f_x=0$ $\quad f_y<0 \quad g_x<0$ 和 $\quad g_y=0$

对于每一种情况，构建合适的相位图，绘出流线，确定均衡的性质。

5 ($a$)证明如果

---

① 斜率与 $m'(p')$ 的绝对值成反比。货币扩张率 $m$ 对通货膨胀率的变化率 $p'$ 的反应越敏感，图 18.4$b$ 中的 $\mu'=0$ 曲线就越平坦。

$f_x<0, f_y>0 \quad g_x<0$ 及 $g_y<0$，那么可能从微分方程组(18.40)或者产生稳定结点均衡，或者是稳定的焦点均衡。

(b)在你构建的相位图中，什么特点对结果(结点或是焦点)的差别起决定性作用？

6 参考奥比斯特模型，验证在图 18.4b 中如果斜率为正的 $\mu'=0$ 曲线足够平坦，那么，尽管流线仍具有交叉的特征，但它们会以结点而非焦点的方式收敛于均衡。

---

## 18.6 非线性微分方程组的线性化

分析非线性微分方程组的另一个定性方法是从该方程组的线性近似中作出推断，这种线性近似是由给定方程组围绕其均衡的泰勒展开推导出来的。① 在 9.5 节中我们知道，对任意函数 $\phi(x)$ 的线性(或者一个高阶多项式)近似可以在展开点给出 $\phi(x)$ 的精确值，但当我们离展开点越来越远时，就会出现逐渐增大的近似误差。对非线性方程组的线性近似也是一样的。在展开点(这里是均衡点 $E$)，线性近似可像原来的非线性方程组一样，精确地确定同样的均衡。在 $E$ 的足够小的邻域内，线性近似应与原方程组具有相同的一般流线图形。因此，只要我们愿意将我们的稳定性推断限定在均衡的紧密邻域内，那么，线性近似可成为一个充足的信息源。这种分析称为局部稳定性分析，它既可以独立运用，也可以运用于相位图分析。这里仅涉及两变量的情况。

639

---

① 在存在多重均衡的情况下，每个均衡需要单独的线性近似。

## 泰勒展开与线性化

给定一个任意(连续可微)单变量函数 $\phi(x)$,围绕点 $x_0$ 的泰勒展开给出级数

$$\phi(x)=\phi(x_0)+\phi'(x_0)(x-x_0)+\frac{\phi''(x_0)}{2!}(x-x_0)^2+\cdots$$
$$+\frac{\phi^{(n)}(x_0)}{n!}(x-x_0)^n+R_n$$

其中包含$(x-x_0)$的各次幂的多项式出现在等号右边。双变量函数 $f(x,y)$围绕$(x_0,y_0)$的泰勒展开式也具有类似的特征。但是在存在两个变量时,所产生的多项式将包含$(y-y_0)$及$(x-x_0)$的各次幂,事实上,是这两个表达式的积:

$$\begin{aligned}f(x,y)=&f(x_0,y_0)+f_x(x_0,y_0)(x-x_0)+f_y(x_0,y_0)(y-y_0)\\&+\frac{1}{2!}\Big[f_{xx}(x_0,y_0)(x-x_0)^2+2f_{xy}(x_0,y_0)(x-x_0)\\&\times(y-y_0)+f_{yy}(x_0,y_0)(y-y_0)^2\Big]+\cdots+R_n\end{aligned} \tag{18.60}$$

注意,现在$(x-x_0)$和$(y-y_0)$的系数是均在展开点$(x_0,y_0)$计值的 $f$ 的偏导数。

由函数的泰勒级数,舍弃掉阶数高于 1 的各项,就可以得到线性近似(或简称为**线性化**)。因此,对于单变量的情况,线性化是 $x$ 的如下线性函数:

$$\phi(x_0)+\phi'(x_0)(x-x_0)$$

类似地,(18.60)的线性化是 $x$ 和 $y$ 的如下线性函数:

$$f(x_0,y_0)+f_x(x_0,y_0)(x-x_0)+f_y(x_0,y_0)(y-y_0)$$

640 此外,用函数符号 $g$ 代替此结果中的 $f$,我们还可以得到对应的

$g(x,y)$的线性化。由此可知,给定非线性方程组

$$x' = f(x,y)$$
$$y' = g(x,y) \tag{18.61}$$

其围绕展开点$(x_0,y_0)$的线性化可以写成

$$x' = f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)$$
$$y' = g(x_0,y_0) + g_x(x_0,y_0)(x-x_0) + g_y(x_0,y_0)(y-y_0) \tag{18.62}$$

如果函数 $f$ 和 $g$ 的具体形式为已知,那么 $f(x_0,y_0)$, $f_x(x_0,y_0)$, $f_y(x_0,y_0)$及 $g(x_0,y_0)$, $g_x(x_0,y_0)$, $g_y(x_0,y_0)$均可以被赋予具体值,而且线性方程组(18.62)可以定量解出。但是,只要 $f_x$, $f_y$, $g_x$, $g_y$ 的符号是可以确定的,那么,即使 $f$ 函数和 $g$ 函数是以一般形式给出的,仍可以对其进行定性分析。

## 约简线性化

根据局部稳定性分析的目的,线性化(18.62)可以简化为更简单的形式。首先,因为展开点将是均衡点$(\bar{x},\bar{y})$,所以我们应以$(\bar{x},\bar{y})$代替$(x_0,y_0)$。而且更根本地,因为由定义,在均衡点我们有 $x'=y'=0$,由此得

$$f(\bar{x},\bar{y}) = g(\bar{x},\bar{y}) = 0 \quad [\text{由}(18.61)]$$

从而在(18.62)每个方程中等号右边的第一项可以省略掉。进行这种变化后,再乘开(18.62)右边余下各项并重排,我们得到线性化的另一种形式:

$$x' - f_x(\bar{x},\bar{y})x - f_y(\bar{x},\bar{y})y = -f_x(\bar{x},\bar{y})\bar{x} - f_y(\bar{x},\bar{y})\bar{y}$$
$$y' - g_x(\bar{x},\bar{y})x - g_y(\bar{x},\bar{y})y = -g_x(\bar{x},\bar{y})\bar{x} - g_y(\bar{x},\bar{y})\bar{y} \tag{18.63}$$

注意,在(18.63)中,等号右边的每一项表示一个常数。我们不辞辛苦地分离出这些常数项是为了可以将其全部省略,从而得到约简的线性化方程。这个结果可用矩阵符号写成

$$\begin{bmatrix} x' \\ y' \end{bmatrix} - \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}_{(\bar{x},\bar{y})} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad (18.64)$$

它构成了(18.61)的约简线性化。由于定性分析仅依赖于对特征根的了解,而对特征根的了解又仅依赖于简化方程组,所以(18.64)足以满足我们局部稳定性分析的全部需要。

再进一步,可以观测到,只有约简线性化的识别的性质存在于在均衡$(\bar{x},\bar{y})$处计值的偏导数矩阵(非线性方程组的雅可比矩阵)641 中。因此,在最终的分析中,均衡的局部稳定性或非稳定性仅是根据上面提到的雅可比矩阵的结构而预测出来的。在后面的讨论中为了符号上的方便,我们以 $J_E$ 表示在均衡点计值的雅可比矩阵,以 $a, b, c, d$ 表示其元素:

$$J_E \equiv \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}_{(\bar{x},\bar{y})} \equiv \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad (18.65)$$

我们假定两个微分方程在函数上不相关,这样我们总有 $|J_E| \neq 0$。(对于 $|J_E| = 0$ 的某些情况,请参见练习 18.6-4。)

## 局部稳定性分析

按照(18.16)并运用(18.65),约简线性化的特征方程应为

$$\begin{vmatrix} r-a & -b \\ -c & r-d \end{vmatrix} = r^2 - (a+d)r + (ad-bc) = 0$$

显然,特征根关键取决于式$(a+d)$和$(ad-bc)$。后者只不过是雅可比矩阵(18.65)的行列式:

$$ad - bc = |J_E|$$

而前者表示雅可比矩阵主对角元素的和,被称作 $J_E$ 的迹(*trace*),以符号 $trJ_E$ 表示:

$$a + d = \mathrm{tr}J_E$$

相应地,特征根可以表示成

$$r_1, r_2 = \frac{\mathrm{tr}J_E \pm \sqrt{(\mathrm{tr}J_E)^2 - 4|J_E|}}{2}$$

$(\mathrm{tr}J_E)^2$ 与 $4|J_E|$ 的相对大小将决定两个根是实根还是复根,即 $x$ 和 $y$ 的时间路径是稳定的还是波动的。另一方面,为检验均衡的动态稳定性,我们需要确定两个根的代数符号。如下两个关系对实现这个目的是极有价值的:

$$r_1 + r_2 = \mathrm{tr}J_E \tag{18.66}$$

$$r_1 r_2 = |J_E| \qquad [\text{参见}(15.5)\text{和}(15.6)] \tag{18.67}$$

**第一种情况** $(\mathrm{tr}J_E)^2 > 4|J_E|$。在此情况下,根为不同实根,不可能出现波动。因此均衡或者是结点均衡,或者是鞍点均衡,但不可能是焦点或涡旋均衡。考虑到 $r_1 \neq r_2$,存在三种不同的符号组合的可能性:两根均为负、两根均为正、两根的符号相反。[①] 考 642 虑到(18.66)和(18.67)中的信息,这三种可能性具有如下特征:

$$(i) \quad r_1<0, r_2<0 \Rightarrow |J_E|>0; \mathrm{tr}J_E<0$$

$$(ii) \quad r_1>0, r_2>0 \Rightarrow |J_E|>0; \mathrm{tr}J_E>0$$

$$(iii) \quad r_1>0, r_2<0 \Rightarrow |J_E|<0; \mathrm{tr}J_E \gtreqless 0$$

在可能性 $i$ 情况下,两个根均为负,当 $t$ 趋于无穷大时,余函数 $x_c$

---

① 因我们排除了 $|J_E|=0$ 的情况,所以没有根取零值。

和 $y_c$ 均趋向于零。因此均衡为稳定的结点。可能性 $ii$ 恰好与 $i$ 相反,它描述了一种非稳定的结点。在两根符号相反的可能性 $iii$ 情况下则产生了鞍点均衡。

为更清楚地了解最后一种可能性,我们回顾一下,在第一种情况下两变量的余函数取如下一般形式

$$x_c = A_1 e^{r_1 t} + A_2 e^{r_2 t}$$

$$y_c = k_1 A_1 e^{r_1 t} + k_2 A_2 e^{r_2 t}$$

其中任意常数 $A_1$ 和 $A_2$ 有待于根据初始条件来确定。如果初始条件使得 $A_1=0$,那么正根 $r_1$ 会被舍弃掉,留下负根 $r_2$ 会使均衡稳定。这样的初始条件从属于位于鞍点稳定枝上的点。另一方面,如果初始条件使得 $A_2=0$,负根 $r_2$ 将消失,留下正根 $r_1$,使得均衡不稳定。这样的初始条件则与位于非稳定枝上的点有关。由于其它的初始条件还包括 $A_1 \neq 0$,它们必然也会产生发散的余函数。因此可能性 $iii$ 会产生鞍点均衡。

**第二种情况** $(\mathrm{tr}J_E)^2 = 4|J_E|$。因为在此情况下的根为重根,所以仅可能发现两种符号组合的可能性:

$$(iv) \quad r_1 < 0, r_2 < 0 \quad \Rightarrow \quad |J_E| > 0; \mathrm{tr}J_E < 0$$

$$(v) \quad r_1 > 0, r_2 > 0 \quad \Rightarrow \quad |J_E| > 0; \mathrm{tr}J_E > 0$$

这两种可能性只是可能性 $i$ 和 $ii$ 的翻版。因此,它们分别指向稳定的结点和非稳定的结点。

**第三种情况** $(\mathrm{tr}J_E)^2 < 4|J_E|$。这种情况下有复根 $h \pm vi$,存在周期性波动,而且我们必定是或者遇到焦点均衡,或者遇到涡旋均衡。在(18.66)和(18.67)基础上,在现在情况下我们有

$$\mathrm{tr}J_E = r_1 + r_2 = (h + vi) + (h - vi) = 2h$$

$$|J_E| = r_1 r_2 = (h + vi)(h - vi) = h^2 + v^2$$

因此，$\mathrm{tr}J_E$ 必然与 $h$ 符号相同，而 $|J_E|$ 必然恒为正。结果会出现三 643
种可能的结果：

$$(vi)\quad h<0 \quad \Rightarrow \quad |J_E|>0;\mathrm{tr}J_E<0$$

$$(vii)\quad h>0 \quad \Rightarrow \quad |J_E|>0;\mathrm{tr}J_E>0$$

$$(viii)\quad h=0 \quad \Rightarrow \quad |J_E|>0;\mathrm{tr}J_E=0$$

它们分别与衰减波动、放大波动以及均匀波动相联系。换言之，可能性 *vi* 意味着稳定的焦点；可能性 *vii* 意味着非稳定的焦点；而可能性 *viii* 则意味着涡旋。

为便于根据 $|J_E|$ 和 $\mathrm{tr}J_E$ 的符号进行定性推断，我们把上面讨论的结论概括在表 18.1 中。该表有三个特点值得注意。首先，负的 $|J_E|$ 仅与鞍点均衡相联系，所以我们可以将 $|J_E|<0$ 作为鞍点均衡的充分必要条件。其次，$\mathrm{tr}J_E$ 仅在两种情况下出现零值——即在鞍点或涡旋的情况下出现零值。而这两种情况根据 $|J_E|$ 的符号是可区分的。相应地，$\mathrm{tr}J_E$ 为零且 $|J_E|$ 为正，是涡旋的充分必要条件。第三，尽管 $\mathrm{tr}J_E$ 为负是动态稳定性的必要条件，但考虑到鞍点均衡的可能性，它还不是充分条件。然而，$\mathrm{tr}J_E$ 为负且有 $|J_E|$ 为正，则是动态稳定性的充分必要条件。

对非线性方程组的线性近似所进行的有关讨论都已概括在表 18.1 中了。但是，表 18.1 中的内容显然也可以应用于起初便是线性的方程组的定性分析。在后一种情况下，雅可比矩阵的元素是一组已知常数，因而没有必要在均衡点计算其值。因为不涉及近似过程，所以稳定性推断本质上不再是局部的，而是具有整体的有效性。

表 18.1　双变量非线性微分方程组的局部稳定性分析

| 情况 | $\|J_E\|$的符号 | $\mathrm{tr}J_E$ 的符号 | 均衡的类型 |
|---|---|---|---|
| 1 $(\mathrm{tr}J_E)^2 > 4\|J_E\|$ | + | − | 稳定结点 |
| | + | + | 非稳定结点 |
| | − | +,0,− | 鞍点 |
| 2 $(\mathrm{tr}J_E)^2 = 4\|J_E\|$ | + | − | 稳定结点 |
| | + | + | 非稳定结点 |
| 3 $(\mathrm{tr}J_E)^2 < 4\|J_E\|$ | + | − | 稳定焦点 |
| | + | + | 非稳定焦点 |
| | + | 0 | 涡旋 |



**例 1**　分析非线性方程组

$$\begin{aligned} x' &= f(x,y) = xy - 2 \\ y' &= g(x,y) = 2x - y \end{aligned} \quad (x, y \geqslant 0)$$

的局部稳定性。首先,令 $x' = y' = 0$,并注意到 $x$ 和 $y$ 的非负性,我们在$(\bar{x}, \bar{y}) = (1,2)$求得单一均衡。然后,通过取 $x'$和 $y'$的偏导数,且在 $E$ 点计算其值,我们得到

$$J_E = \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}_{(\bar{x},\bar{y})} = \begin{bmatrix} y & x \\ 2 & -1 \end{bmatrix}_{(1,2)} = \begin{bmatrix} 2 & 1 \\ 2 & -1 \end{bmatrix}$$

因为$|J_E| = -4$ 为负,所以我们可以立刻得出结论,均衡是局部鞍点均衡。

注意,尽管雅可比矩阵的第一行最初包含变量 $x$ 和 $y$,但第二行却不包含 $x$ 和 $y$。存在这种差别的原因在于给定方程组中的第二个方程原来就是线性的,无需进行线性化。

**例 2**　给定非线性方程组

$$x' = x^2 - y$$

$$y' = 1 - y$$

令 $x' = y' = 0$,我们可以求得两个均衡点:$E\ 1 = (1,1)$ 和 $E_2 = (-1, 1)$。因此我们需要两个单独的线性化。在两个均衡点依次计算雅可比矩阵 $\begin{bmatrix} 2x & -1 \\ 0 & -1 \end{bmatrix}$ 的值,我们得到

$$J_{E1} = \begin{bmatrix} 2 & -1 \\ 0 & -1 \end{bmatrix} \quad 和 \quad J_{E2} = \begin{bmatrix} -2 & -1 \\ 0 & -1 \end{bmatrix}$$

第一个矩阵具有一个负的行列式,因此 $E_1 = (1,1)$ 是一个局部的鞍点。由上面第二个矩阵,我们得到 $|J_{E2}| = 2$ 和 $\mathrm{tr}J_{E2} = -3$。因此,根据表 18.1 可知,$E_2 = (-1,1)$ 是第一种情况下的局部稳定结点。

**例 3** 线性方程组

$$x' = x - y + 2$$
$$y' = x + y + 4$$

具有稳定的均衡吗?为回答此定性问题,我们可以仅集中于简化方程,并忽略常数 2 和 4。正如由线性方程组可以预期的那样,雅可比矩阵 $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ 的四个元素均为常数。由于其行列式和迹均等于 2,所以均衡属于第三种情况,是一个不稳定的焦点。注意,得到此结论毋须解得均衡。还要注意,在此情况下的结论是整体有效的。

**例 4** 分析奥比斯特模型(18.50)的局部稳定性, 645

$$p' = h(1 - \mu)$$
$$\mu' = (p + q - m)\mu$$

假定货币扩张率 $m$ 是外生的(不遵从货币规则)。按照图 18.4$a$,此模型的均衡发生于点 $E = (\bar{P}, \bar{\mu}) = (m - q, 1)$。在 $E$ 点计值的

雅可比矩阵为

$$J_E = \begin{bmatrix} \frac{\partial p'}{\partial p} & \frac{\partial p'}{\partial \mu} \\ \frac{\partial \mu'}{\partial p} & \frac{\partial \mu'}{\partial \mu} \end{bmatrix}_E = \begin{bmatrix} 0 & -h \\ \mu & p+q-m \end{bmatrix}_{(m-q,1)} = \begin{bmatrix} 0 & -h \\ 1 & 0 \end{bmatrix}$$

因为$|J_E| = h > 0$,$\mathrm{tr}J_E = 0$,表 18.1 表明,均衡是一个局部的涡旋均衡。这个结论与上一节相位图分析所得出的结论是一致的。

**例 5** 假设可供选择的货币规则如下:

$$p' = h(1-\mu) \qquad [\text{源于}(18.50)]$$

$$\mu' = [p + q - m(p')]\mu \qquad [\text{源于}(18.58)]$$

分析奥比斯特模型的局部稳定性。注意,因 $p'$ 是 $\mu$ 的函数,所以在现在的模型中,函数 $m(p')$ 也是 $\mu$ 的函数。在均衡点 $E$ 处,$p' = \mu' = 0$,我们有 $\bar{\mu} = 1$ 和 $\bar{p} = m(0) - q$。因此,在 $E$ 点计值的雅可比矩阵为

$$J_E = \begin{bmatrix} 0 & -h \\ \mu & p+q-m(p')-m'(p')(-h)\mu \end{bmatrix}_E$$

$$= \begin{bmatrix} 0 & -h \\ 1 & m'(0)h \end{bmatrix}$$

其中由(18.57)知,$m'(0)$为负。按照表 18.1,在具有$|J_E| = h > 0$ 和 $\mathrm{tr}J_E = m'(0)h < 0$ 时,我们是能得到稳定的焦点还是稳定的结点,要视$(\mathrm{tr}J_E)^2$ 和 $4|J_E|$的相对大小而定。具体而言,导数$m'(0)$的绝对值越大,$\mathrm{tr}J_E$ 的绝对值将越大,$(\mathrm{tr}J_E)^2$ 越有可能超过 $4|J_E|$,从而产生稳定的结点而非稳定的焦点。这个结论与我们前面相位图分析所得到的结论同样是一致的。

## 练习 18.6

1　分析下列每个非线性方程组的局部稳定性：

(a) $x' = e^x - 1$　　(c) $x' = 1 - e^y$

$y' = ye^x$　　$y' = 5x - y$

(b) $x' = x + 2y$　　(d) $x' = x^3 + 3x^2y + y$

$y' = x^2 + y$　　$y' = x(1 + y^2)$

2 给定

646

(a) $f_x = 0$　$f_y > 0$　$g_x > 0$　和　$g_y = 0$

(b) $f_x = 0$　$f_y < 0$　$g_x < 0$　和　$g_y = 0$

(c) $f_x < 0$　$f_y > 0$　$g_x < 0$　和　$g_y < 0$

运用 18.1 确定一个非线性方程组具有的局部匀衡的类型。你的结果与你在练习 18.5－4,18.5－5 所得到的结果一致吗？

3　假设遵守传统的货币规则，分析奥比斯特模型的局部稳定性。

4　如下两个方程组均具有零值的雅可比行列式。构建每一方程组的相位图,并推断存在的所有均衡的位置：

(a) $x' = x + y$　　(b) $x' = 0$

$y' = -x - y$　　$y' = 0$

# 18.7　动态分析的局限性

本书第二篇所介绍的静态分析仅讨论这样的问题：在模型给定的条件下，如何确定均衡的位置？主要的问题是：如果达到均衡，那么什么值是变量恒久不变的均衡值呢？但均衡位置的可达到性被视为理所当然的。当我们进入第三篇比较静态学的领域时，关注的中心转移到另一个更有趣的问题：均衡位置如何随参数

的某些变化而移动？但均衡的可达到性问题依然被撇在一边。直到第五篇我们进入动态分析的王国时，我们才正视均衡的可达到性问题。在这里，我们具体地问这样的问题：如果我们最初偏离均衡位置（比如，由于打破均衡的参数的新变化），模型中的各种力量会驱使我们走向新的均衡吗？而且，在动态分析中，我们还能了解变量趋向于均衡（如果它存在的话）的路径的特征（是稳定的、波动的还是振荡的），因此，动态分析的重要性是不言而喻的。

但是在结束对动态分析的讨论的时候，我们还需要认识到动态分析的局限性。首先，为使分析易于把握，动态模型一般是按线性方程来构建的。尽管这样可以获得简洁性，但线性假定在许多情况下会极大地损害模型的现实性。因为与线性模型有密切关系的时间路径并不总是近似于对应的非线性模型的时间路径，正如我们在16.6节价格上限例子中所看到的那样，所以，在解释和应用线性动态模型的结果时，必须保持足够的谨慎。但在这方面，定647性图示法可能会发挥更大的作用，因为这种方法可在非常一般的条件下把非线性纳入模型，却没有使分析增加更多的复杂性。

动态经济模型另一个常见的缺点是在微分或差分方程中使用常系数。由于系数的主要作用是设定模型的参数，所以，系数的不变性（这也是为了数学上的可处理性）实质上是“冻结了”所研究问题的经济环境。换言之，这意味着模型的内生调整是在某种经济真空中进行研究的，以致不允许外生因素进行干扰。当然，在某些情况下，这一问题并不十分严重，因为许多经济参数确实在较长的时期内相对稳定。在另一些情况下，我们可以进行比较动态分析，以考察变量的时间路径如何被某些参数的变化所影响。然而，在解释趋向于遥远未来的时间路径时，如果模型进行了简化的常系

数假设，那么，我们一定要非常谨慎，对时间路径在更远的时期中的有效性不能过于自信。

当然，读者应当认识到，我们这里指出动态分析的局限性，决不是为了贬低动态分析。事实上，大家还记得，对于迄今所介绍的每一种分析方法，我们都介绍了其局限性的一面。因此，只要动态分析能够被恰当地解释和适当地应用(同其它类型的分析一样)，那么，它就能在研究经济现象中发挥重要作用。



# 第六篇

# 数　学　规　划

651

# 第19章 线性规划

在本书前五篇,我们已经研究了静态学、比较静态学、动态学整个分析体系。其中在静态学或均衡分析中研究了几类最优化问题。但从方法论上看,我们还局限于主要以微积分为基础的古典最优化方法。在本篇,我们将介绍一种最新发展的被称为数学规划(包括线性规划和非线性规划)的方法。

数学规划与古典最优化的不同之处在于:它力图解决最优化寻求者面对不等式约束时的优化问题,比如这些约束可能以$g(x,y)\leqslant c$,而不是以$g(x,y)=c$的形式出现。举例来说,这种数学规划的框架允许一个消费者自由选择是花掉250美元,还是少于250美元,而不是要求他恰好花掉250美元。因此,通过放宽约束条件,这种新的最优化框架与上述问题变得更有意义,而且更加现实。但它也要求发展出一种新的求解方法,因为不等式约束问题难以用古典的微积分方法来处理。

我们将从一种比较简单的规划问题——线性规划开始,其目标函数及约束条件均是线性的。

## 19.1 线性规划的简单示例

652 线性规划的本质内容可以通过具体例子很好地表达出来。我们将举两个例子,一个用于说明最小化问题,另一个用于说明最大

化问题。

## 饮食问题

为保持健康,一个人每日摄入的几种营养成份必须达到某一最低量。为简化计,假设只考虑三种营养成份:钙、蛋白质及维生素 A。我们还假设此人的饮食仅由两种食物构成,即食品 I 和食品 II,其价格和营养成份列在表 19.1 中,我们还在表中列出了每日所需的各种营养物质的最小量。问题是:两种食物的何种组合能够既满足日常需要,又能使成本最低呢?

若我们以 $x_1$, $x_2$ 表示每日购买的两种食物的数量(将 $x_1$, $x_2$ 视为连续变量),此问题在数学上可表述如下:

$$
\begin{aligned}
&\text{Min. } C = 0.6x_1 + x_2 \\
&\text{s.t.} \quad 10x_1 + 4x_2 \geqslant 20 \quad [\text{钙约束条件}] \\
&\qquad\quad 5x_1 + 5x_2 \geqslant 20 \quad [\text{蛋白质约束条件}] \\
&\qquad\quad 2x_1 + 6x_2 \geqslant 12 \quad [\text{维生素A约束条件}] \qquad (19.1) \\
&\text{并且:} \quad x_1, x_2 \geqslant 0
\end{aligned}
$$

**表 19.1　食品的价格及其营养成分**

| 价格 | 食品 I(每磅) | 食品 II(每磅) | |
|---|---|---|---|
| | 0.60 美元 | 1.00 美元 | 每日所需最少量 |
| 钙(单位*) | 10 | 4 | 20 |
| 蛋白质(单位*) | 5 | 5 | 20 |
| 维生素 A(单位*) | 2 | 6 | 12 |

* 表示一种假想的单位,为的是在例子中能方便地取整数。

(19.1)中的第一个方程,是一个以表 19.1 中的价格信息为基础的成本函数,它构成了线性规划的目标函数;随后的三个不等式是由

日常营养需要组成的约束条件，它们可以很容易地从表中最后三行数据变换而来。读者应注意到，虽然不允许低于日常营养需要，但由于这里使用的是弱不等号≥，所以最优化寻求者可以超过表中所列的最低量。正是这一特点，构成了线性规划与前面讨论的最优化问题的主要区别。最后，我们利用两个所谓的非负限制的附加不等式 $x_1,\ x_2 \geq 0$，把由于微积分的局限性而隐含于古典最优化框架中的要求，即“购买量不可能为负”明确表示出来。还应653注意，在我们的问题中，约束条件的个数实际上多于选择变量的个数，这在古典最优化中是从未出现过的，但在这里却是行得通的，因为约束条件已由等式弱化为不等式，因此它更易于得到满足。

简言之，线性规划中存在三种基本成分：目标函数、一组约束条件以及一组非负性限制。注意，这里所讨论的问题都是线性的，因为没有一个变量的幂数不是 1，也没有一个变量与另一个变量相乘。当然，正是由于这一事实，才产生线性规划这一名称。

## 图解法

因为我们的问题中只包括两个变量，所以它是可以用图形分析的方法求解的。在图 19.1 中，两坐标轴分别表示 $x_1$ 和 $x_2$。由于非负限制，我们只需考虑非负象限。

为了观察约束条件所表示的曲线的形状，我们首先假定约束条件是以三个方程的形式给出的，并将其绘成如图 19.1 $a$ 中所示的三条直线。它们中的每一条分别表示钙边界、蛋白质边界、维生素A边界，并将非负象限分成两个非重叠区域。因为每个约束条件都是“≥”类的，所以只有那些位于边界上或边界东北区域内的点（有序偶）能够满足有关的特定约束条件。为同时满足所有三个

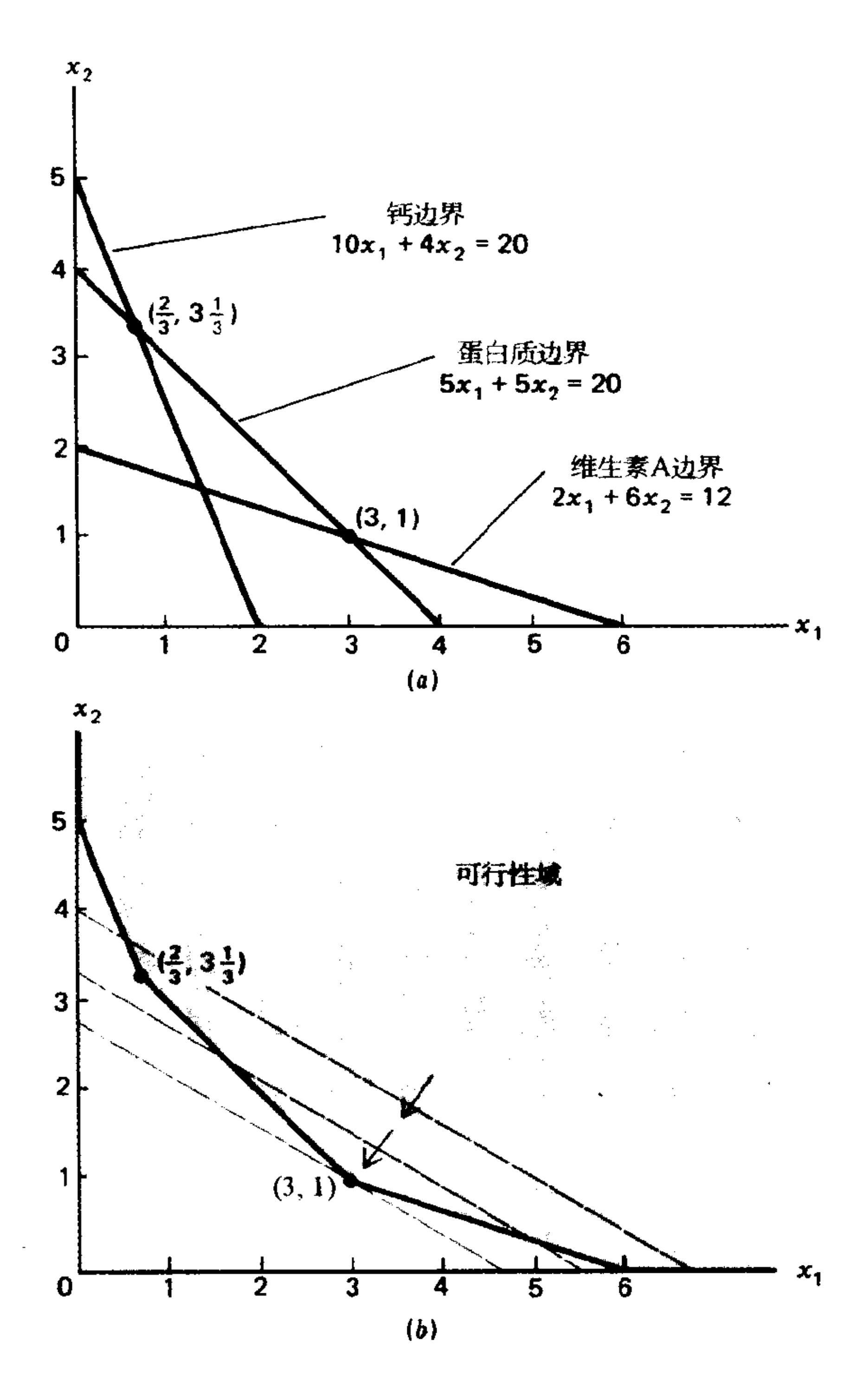



图 19.1

约束条件,我们只能接受这样的有序偶$(x_1, x_2)$,它们不位于我们所绘任一条边界的西南部。例如点(1,2)的确能满足维生素 A 的约束条件。但它不满足另两个约束条件。因此,在我们的线性规

划中,它是不可行的。而所有位于图 $b$ 中阴影区内的点的确都能同时满足三个约束条件。为此,该阴影区域被称作可行区域,而在这个区域中的每个点(有序偶)也被称作一个可行解。读者应知道,可行区域包括位于边界交界处的那些点(实点),特别是,水平轴上的点的集合 $\{(x_1, x_2) \mid x_1 \geqslant 6, x_2 = 0\}$ 和纵轴上点的集合 $\{(x_1, x_2) \mid x_1 = 0, x_2 \geqslant 5\}$,也都属于可行区域。因此,可以将可行区域看成是一个(有序偶)的闭集。闭集是一个与闭区间相似的概念,它是一个包括所有边界点的集合。

应注意,可行区域的弯折边界是由三个约束边界所选分段构654成。还应注意[在我们现在两维情况下]边界上的角点——称作极点——分别位于两边界线的交点[例如(3,1)和($\frac{2}{3}$,3 $\frac{1}{3}$)或一个边界线与一个坐标轴的交点[例如(0,5)和(6,0)]。这些极点在我们今后的求解中将会证明是很有意义的。

位于可行区域中的点表示所有既能满足所有约束条件又能满足非负限制的点的集合,但是其中一些还必须使购买成本相对较655小。为使成本 $C$ 最小,还必须考虑到目标函数。成本函数可写成:

$$x_2 = C - 0.6x_1$$

如果我们将 $C$ 看成是一个参数,那么我们就可以将此方程绘成具有相同斜率 $-0.6$ 的一族平行线。其中,在图 19.1$b$ 中用虚线表示出了三条可能的直线。由于这些线中的每一条都对应于某一确定的 $C$ 值,我们所绘制的实际上是三条等成本线。为了使成本最小,我们必须挑选那条尽可能最低的且位于可行区域中的等成本线。在图 19.1$b$ 中,我们挑选了极点(3,1)。因此,我们线性规划

中的最优可行解(或简言之,最优解)就是$(\bar{x}_1, \bar{x}_2)=(3,1)$。随后,根据所给出的食品价格计算,可以得到该饮食的最小成本为每天2.80美元。假想的最优化者每天吃3磅食品I和1磅食品II,这种一成不变的吃法当然会令人厌烦,但就生活的成本而言,这种组合当然是最好的。

根据以上所描述的,饮食问题不过是最低成本组合问题的一种变形而已。等成本概念与以前完全一致,而以前所遇到的平滑等量曲线现在换成了具有弯折边界的可行区域。因此,微积分中切点的概念必须摒弃而代之以接点的概念。在图19.1$b$中,接点是位于边界上的一个极点。应指出的是,这不仅仅是一种巧合。正如我们以下将会看到的,所有线性规划的最优解总可以在极点处得到。因为极点总是表示两条约束边界(或一个约束边界与一个坐标轴)的交点,我们可以在发现那个最佳隅角后,通过解两个相交直线的方程便可准确地求出最优解$(\bar{x}_1, \bar{x}_2)$。在我们现在这个例子中,那个最佳隅角位于蛋白质边界与维生素A边界的相交处。因此,通过解以下两个方程:

$$5x_1 + 5x_2 = 20$$

$$2x_1 + 6x_2 = 12$$

我们可以得出$(\bar{x}_1, \bar{x}_2)=(3,1)$。

可以看出,这个解完全满足蛋白质和维生素$A$的需求量,但它超出了所需钙的量。如果所有的约束条件都是严格的方程,那么这个解是不可能出现的。

## 价格变化的影响

我们现在提出一个比较静态问题:如果两种食品的价格$P_1$

和 $P_2$ 发生了变化，那么最优解将会如何呢？因为等成本线的斜率由 $-P_1/P_2$ 的比率来度量的（在我们的例子中，$-0.60/1.00=-0.6$），这种价格变化的影响立即就可以反映到那些等成本线上，但这也存在几种可能性。

656 首先，如果两种价格以完全相同的比例变化，那么等成本线的斜率将不变。在这种情况下，虽然平均成本值 $\bar{C}$ 将随 $P_1$ 和 $P_2$ 的变化而很自然地上升或下降，但是初始的最优解一定仍然继续适用。

其次，两种价格以不同的比例变化，但变化的量相对较小。在这种情况下，等成本线的斜率将发生很小的变化，例如从 $-0.6$ 到 $-0.4$ 或到 $-0.8$。读者可以通过在图 19.1$b$ 中绘出的一族具有 $-0.8$ 斜率的等成本线来证实这一点。斜率这种数量上的变化将对原始的最优解没有影响。因此，不像微积分中的切点，接点（最佳隅角）对价格参数中的微小变化并不敏感。

还有另外一种可能性。假定现在两种价格相等，如 $P_1=P_2=1$。那么现在由于等成本线的斜率为 $-1$，它将平行于蛋白质的约束边界。新的，可能最低的等成本线与可行区域相接触处不是一个单独的点而是其约束边界的一条完整的边。因此，这条线上从 $(3,1)$ 到 $(\frac{2}{3},3\frac{1}{3})$ 上的每一点均为最优解。就最优化者而言，这种多重最优现象的产生是很正常的。相反，这也可以算作一件幸运的事情，因为现在可能会使菜谱中有一些变化。但是，对我们而言，这种现象好像要促使我们收回我们以前说过的话，也就是最优解总是位于极点处。但是，稍微思考一下，我们仍然站得稳脚，因为即使是在存在多重最优解的情况下，最优解仍位于一个

隅角处——甚至是两个隅角处！事实上，如果我们仅仅将注意力集中到极点，我们就会冒失去较好的解的风险。我们将发现，正是这条思路构成了我们下边将要介绍的所谓“单形法”的理论基础。

## 生产问题

让我们接下来再转到一个简单的例子,这次涉及到的是生产问题。

假设如下:一个厂商生产两个系列产品,产品 I 和产品 II,此工厂由三个生产车间构成,即切割车间、混合车间和包装车间。每个车间的设备每天可使用 8 个小时;因此,我们可将 8 小时看成每个车间每天的工作能力。产品的生产过程可概述如下:(1)产品 I 首先切割,然后包装。每吨该产品用掉切割能力的 1/2 小时和包装能力的 1/3 小时。(2)产品 II 首先混合,然后包装。每吨该产品用掉混合能力的 1 小时和包装能力 2/3 小时。最后,产品 I 和产品 II 分别以每吨 80 美元和 60 美元的价格出售,但是分别扣除发生的可变成本后,每吨产品分别可以产生 40 美元和 30 美元的净值。这些数字既可以看作是每年的净收益(不含可变成本),也可以看作是毛利润数(包含固定成本)。为简化计,我们这里将它们称为“每吨的利润”。问题是:为了使总(毛)的利润最大,该厂商应选择怎样的产量组合?

为便于回答这个问题,我们先将已知信息列成表的形式,见表 19.2,然后将该问题转化成用两个连续的选择变量 $x_1$ 和 $x_2$ 表 657
示的线性规划问题:

$$\text{Max.} \qquad \pi = 40x_1 + 30x_2$$

$$
\begin{aligned}
\text{s.t.} \quad & x_1 \leqslant 16 & [\text{切割约束条件}] \\
& x_2 \leqslant 8 & [\text{混合约束条件}] \\
(19.2) \quad & x_1 + 2x_2 \leqslant 24 & [\text{包装约束条件}] \\
\text{且：} \quad & x_1, x_2 \geqslant 0
\end{aligned}
$$

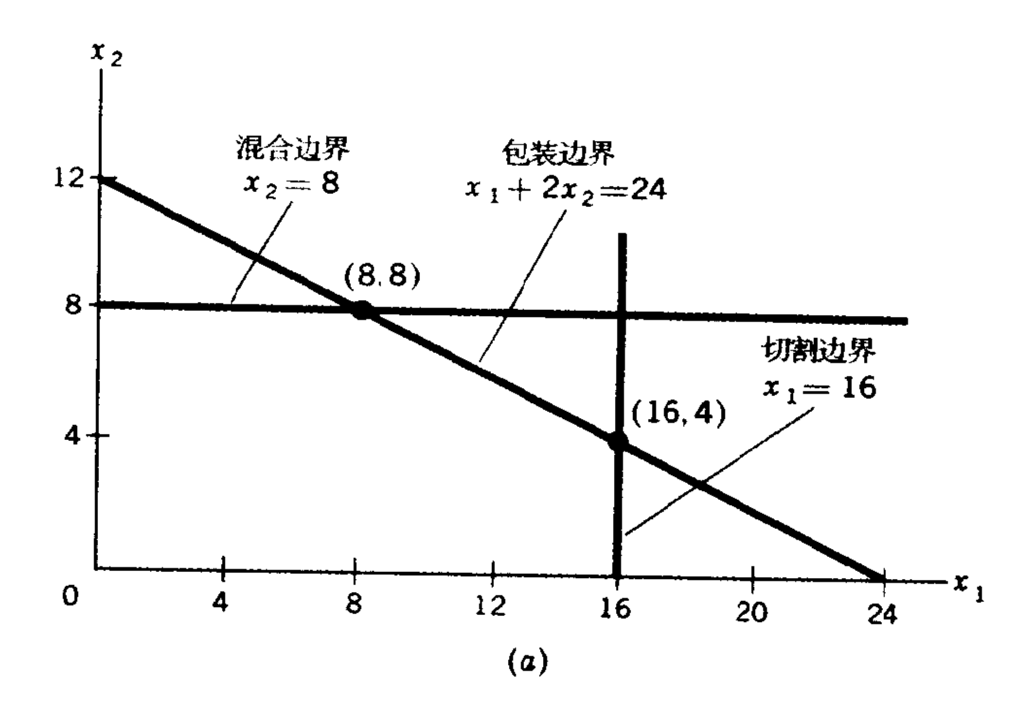
注意，虽然按照表 19.2，切割的约束条件应是$\frac{1}{2}x_1 \leqslant 8$，但是，为了不用分数表示，我们已将其两边同乘以 2。类似地，我们也已用 3 去乘包装的约束条件而对其进行变换。现在该问题就是一个最大化的问题。还有，现在约束条件是"≤"类型的，这是因为即使我们不能超出生产能力的限度，但是我们有权使其部分闭置。而非负限制仍像在最小化问题那样，以同样的形式出现。

## 图解法

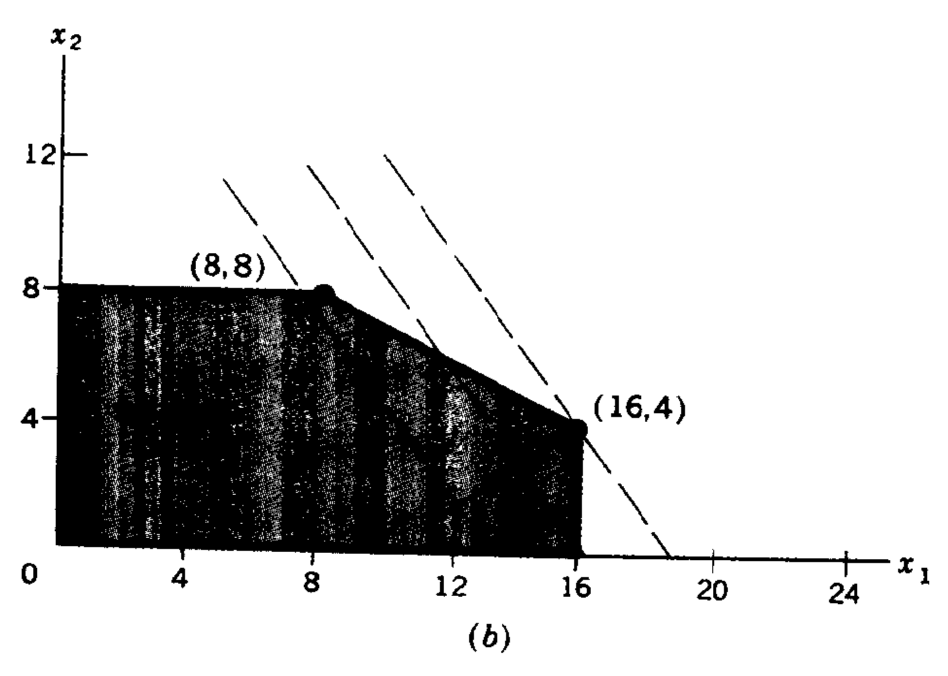
(19.2)中的线性规划同样可以像在图 19.2 中那样用图解法来解。由于非负限制，这个问题又一次被限制在非负象限。在该象限中，我们绘制了三条约束边界。切割边界($x_1=16$)和混合边界($x_2=8$)，如图 19.2*a* 中所示，分别为一垂直与水平直线，而包装约束边界则是一条斜线，分别与另两条约束边界交于点(16,4)和(8,8)。

因为我们的约束条件是"≤"类型的，所以这次的可行区域将由同时满足以下三个位置设定条件的点集构成：(1)位于混合边界上或其下方。(2)位于切割边界上或其左方。(3)位于包装边界上或其下方。我们可以将所有满足这些条件的点所构成的集合用图 19.2*b* 中的阴影部分表示。因为这个区域的所有边界点及其内点都是可行解，所以该可行区域是一个闭集。在现在这种情况下，也

658



(a)



(b)

图 19.2

可称该集合是严格有界的。其意思大致是，这个阴影的区域能够被“装进”一个有确定尺寸的盒子中。（与此相反，图 19.1b 中的可行区域只是下有界。）

**表 19.2　产品利润与产品加工设备**

| | 每吨所需的加工时数 | | |
|---|---|---|---|
| | 产品 I | 产品 II | 日加工能力(时数) |
| 切割 | $\frac{1}{2}$ | 0 | 8 |
| 混合 | 0 | 1 | 8 |
| 包装 | $\frac{1}{3}$ | $\frac{2}{3}$ | 8 |
| 每吨利润 | \$ 40 | \$ 30 | |

最后,我们考察一下目标函数,该函数可改写成:

$$x_2 = \frac{\pi}{30} - \frac{4}{3}x_1$$

若将$\pi$ 看成参数,我们可以将此方程绘成具有共同斜率 $-\frac{4}{3}$ 的一659族直线,在 $b$ 图中可以看到其中的三条(用虚线表示)。因为这些直线中的每一条都对应着利润参数$\pi$的某一特定值,所以我们可以称它们为等利润线。我们的目标当然是在可行区域中尽可能地得到最高的等利润线。因此,我们必须挑选极点(16,4)作为所代表的最佳产量组合。也就是,最优解为每天生产 $\bar{x} = 16$ 吨,$\bar{x}_2 = 4$ 吨。将这些数值代入到目标函数中,我们就可以得出每天的最大毛利润为 760 美元。

注意,解(16,4)完全满足切割和包装的约束条件,但并不完全满足混合的约束条件。为了使利润最大化,一部分混合生产能力应闲置不用。如果这些约束条件都是以严格的方程形式给出的话,那么这样的结果将是不可想象的。

提到约束条件的形式，虽然上述的两个例子中的约束条件要么全是“≤”类型的(最大化问题)，要么全是“≥”类型(最小化问题)，但是它们不必总是如此。可以想象：例如，对饮食中维生素A的摄取量，可以通过双重约束条件“不少于12”(为了健康)和“不超过86”(为了避免过度累积)，这样就会出现两种不等式。但是，这并不影响解法，因为在(19.2)中，我们已经看到非负限制的“≥”类型和生产能力的约束的“≤”类型同时共存的情形。当两类约束条件都存在时，确实会产生约束条件不一致的可能。在图19.3*a*的情况下，如果该问题要求可行区域既位于边界I上方，又位于边界II下方，那么可行区域将是一个空集，因而这个问题不能求解。

一个线性规划实际上也可能包括一个或多个**等式约束**。图19.3*b*所表示的(阴影)可行区域与图19.2*b*中的相似，但是，若附

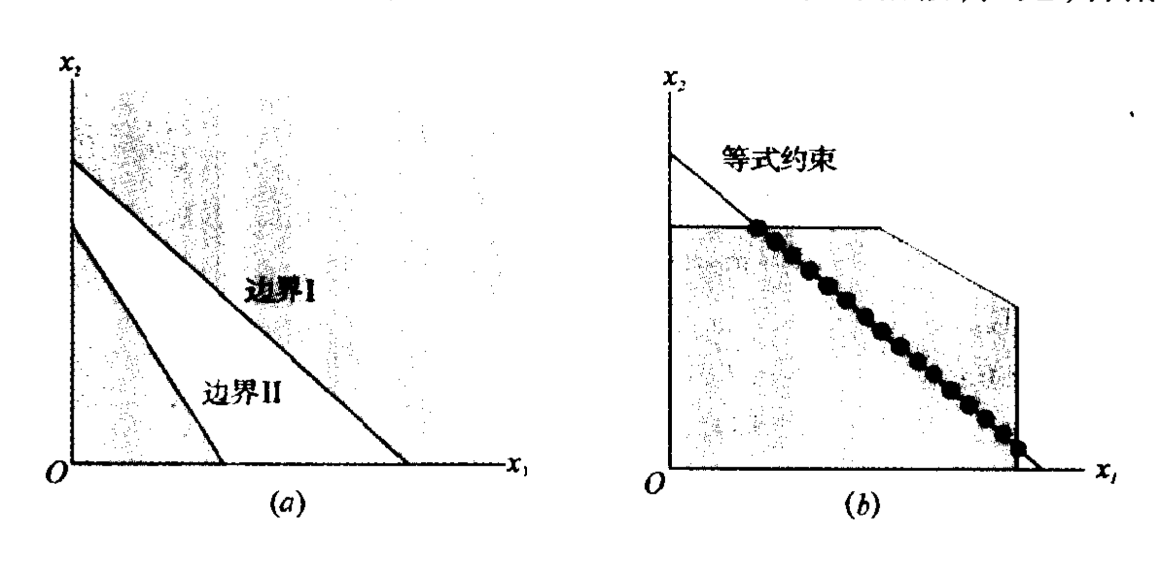



图 19.3

加上一个等式约束，那么可行解集就从一个阴影区域收缩成打上圆点的线段——阴影区域与新的约束直线的交集。不过，同样的求解方法仍会适用。

 **练习 19.1**

1 用图解法解下列各题：

(*a*) Max. $\pi = 2x_1 + 5x_2$

$$\text{s.t.}\quad \begin{aligned} x_1 \quad &\leqslant 4 \\ x_2 &\leqslant 3 \\ x_1 + 2x_2 &\leqslant 8 \end{aligned}$$

且：$x_1,\ x_2 \geqslant 0$

(*b*) Min. $C = 12x_1 + 42x_2$

$$\text{s.t.}\quad \begin{aligned} x_1 + 2x_2 &\geqslant 3 \\ x_1 + 4x_2 &\geqslant 4 \\ 3x_1 + x_2 &\geqslant 3 \end{aligned}$$

且：$x_1,\ x_2 \geqslant 0$

2 哪些特点使得下列各题的图形有显著区别？这些特点又是怎样产生的？

(*a*) Min. $C = 6x_1 + 24x_2$

$$\text{s.t.}\quad \begin{aligned} x_1 + 2x_2 &\geqslant 3 \\ x_1 + 4x_2 &\geqslant 4 \end{aligned}$$

且：$x_1,\ x_2 \geqslant 0$

(*b*) Min. $C = 6x_1 + 30x_2$

s.t. 与(*a*)中同样的约束。

3 一个房主打算将她的房子粉刷一下。这个房子只需粉刷一层。为了满足这个要求，该油漆必须至少应具有 200 厘泊的粘度。为了达到理想的亮度，另一个要求是，在每加仑油漆中必须至少含有 14g 的化学成分 *Y*，为了达到理想的耐久性，在每加仑油漆中必须至少含有 30g 的另一种化学成分 *Z*。

可供她使用的油漆有两种(I 和 II)，I 的价格为每加仑 6 美元，II 的价格为每加仑 4 美元。它们的规格包含在如下的表格中：

|  | 油漆 I(每加仑) | 油漆 II(每加仑) |
|---|---|---|
| 粘度(厘泊) | 400 | 100 |
| $Y$(克) | 20 | 10 |
| $Z$(克) | 20 | 60 |

为满足上述三个要求且使成本最小,在每加仑中应如何混合 I 和 II 类两种油漆(称这些为 $x_1$ 和 $x_2$)? 该混合油漆的最小成本是多少? (提示:除了那些正常的约束条件外,还应有一个等式的约束条件 $x_1+x_2=1$,为什么?)

4 在前面的问题中,如果将油漆 I 和油漆 II 的价格对调一下,对油漆混合将有什么影响? 每加仑新的混合油漆的最小成本应是多少?

5 如果下式成立的话,那么有两个选择变量的线性规划的图形将会发生怎样的变化?

(*a*) 目标函数中 $x_1$ 的系数为零。

661

(*b*) 目标函数中 $x_2$ 的系数为零。

(*c*) 前两个约束条件分别为:

$$ax_1 + bx_2 \leqslant c$$

$$-ax_1 - bx_2 \leqslant -c$$

6 当地某个公司计划在某一特定星期在广播与电视上做一个特殊销售周年纪念的广告。为此,最大预算为 16000 美元。现了解到,广播广告的成本是每 30 秒为一时间点(称为 $x_1$),价格为 800 美元,且至少需订 5 个时间点的合约。另一方面,电视广告每时间点(称为 $x_2$)价格为 4000 美元。由于再三要求,在特定的星期中也只能利用 4 个电视时间点。以所估计的听众规模及其它因素为基础,确信,在开发潜在消费者方面,每个电视时间点的影响力将是广播的 6 倍。为最大限度地吸引潜在消费者,公司应如何分配其广告?

7 上题中,如果利用电视做广告的时间点没有限制,那么最优解会如何呢?

## 19.2 线性规划的一般表示

上节的两个例子说明了两个选择变量和三个约束条件的情

况。当代之以有 $n$ 个选择变量和 $m$ 个约束条件时，线性规划仍具有与其相同的一般形式——其主要成分是：一个线性目标函数，一组线性的不等式约束条件和一组非负限制。推广到含有 $n$ 个变量的线性规划，可有三种可供选择的表达方法：用普通写法，用求和($\sum$)符号或用矩阵符号。

### 普通写法

含有 $n$ 个变量且满足 $m$ 个约束条件的最大化规划问题完全写出来时，将是如下形式：

$$
\begin{aligned}
&\text{Max.} \quad \pi = c_1x_1 + c_2x_2 + \cdots + c_nx_n \\
&\text{s.t.} \quad a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \leqslant r_1 \\
&\qquad a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \leqslant r_2 \\
&\qquad \vdots \qquad \vdots \qquad\qquad \vdots \qquad \vdots \\
&(19.3) \quad a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \leqslant r_m \\
&\text{且：} \quad x_j \geqslant 0 \ (j = 1, 2, \cdots, n)
\end{aligned}
$$

在(19.3)中，尽管在许多场合中，目标函数与利润函数有所不

662 同，但是，我们还是借用在生产例子中的符号 $\pi$ 作为表示被**求极大函数**(求极大化的对象)的一般性符号。选择变量用 $x_j(j = 1, 2, \cdots, n)$表示，它们在目标函数中的系数为 $c_j(j = 1, 2, \cdots, n)$，它是一组已知的常数。符号 $r_i(i = 1, 2, \cdots, m)$——另外一组常数——另一方面，表示加到规划中的一些“限制”。为了统一化，我们将所有 $m$ 个约束条件都写成“$\leqslant$”的不等式，但这也不失一般性。特别需要注意的是，在一个约束条件以“$\geqslant$”形式出现的情况下，我们总可以通过将不等式两边同乘以 $-1$ 而将其变成“$\leqslant$”的形式。最后，注意在约束条件中，选择变量的系数用 $a_{ij}$ 表示，这里

的双下标可注明每个系数的特定位置。因为一共有 $m$ 个含有 $n$ 个变量的约束条件(这里 $m$ 可以大于,等于或小于 $n$),所以系数 $a_{ij}$ 将构成一个 $m \times n$ 维的矩阵。

类似地,一个最小化规划用普通写法可以写成如下形式:

$$
\begin{aligned}
\text{Min.}\quad & C = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n \\
\text{s.t.}\quad & a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n \geqslant r_1 \\
(19.4)\quad & a_{21} x_1 + a_{22} x_2 + \cdots + a_{2n} x_n \geqslant r_2 \\
& \vdots \qquad \vdots \qquad\qquad \vdots \qquad \vdots \\
& a_{m1} x_1 + a_{m2} x_2 + \cdots + a_{mn} x_n \geqslant r_m \\
\text{且:}\quad & x_j \geqslant 0 \ (j = 1, 2, \cdots, n)
\end{aligned}
$$

即使在许多场合中的目标函数也许不是成本函数,但是我们还是又一次借用饮食问题中的符号 $C$ 作为表示被求极小值函数(求极小值的对象)的一般性符号。目标函数中的 $c_j$ 仍表示已知的一组常系数,像约束条件中的 $r_i$ 一样。但是在现在这个场合,符号 $r$ 表示需求而不是限制。约束条件中的系数和选择变量的符号保持不变。但现在约束条件都以"$\geqslant$"不等式的形式出现。

## 求和符号($\sum$)

将(19.3)和(19.4)中的线性规划用求和符号($\sum$)表示是最能节省空间的:

$$
\begin{aligned}
\text{Max.}\quad & \pi = \sum_{j=1}^{n} c_j x_j \\
\text{s.t.}\quad & \sum_{j=1}^{n} a_{ij} x_j \leqslant r_i \qquad (i = 1, 2, \cdots, m) \\
\text{且:}\quad & x_j \geqslant 0 \qquad (j = 1, 2, \cdots, n)
\end{aligned}
$$

663 类似地，

$$\text{Min.} \qquad C = \sum_{j=1}^{n} c_j x_j$$

$$\text{s.t.} \qquad \sum_{j=1}^{n} a_{ij} x_j \geqslant r_i \qquad (i = 1, 2, \cdots, m)$$

$$\text{且:} \qquad x_j \geqslant 0 \qquad (j = 1, 2, \cdots, n)$$

虽然这种表达方式很简要，但是以"∑"形式表达，在数学操作方面，是很不方便的。因此，我们下面将不用这种表达方式。

## 矩阵符号

为了解矩阵符号是如何应用的，我们先定义一下以下四个矩阵：

$$(19.5) \qquad c \equiv \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \quad x \equiv \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$A \equiv \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad r \equiv \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}$$

其中的三个为列向量：$c$ 和 $x$ 为 $n \times 1$ 维的，$r$ 为 $m \times 1$ 维的。矩阵 $A$ 是一个 $m \times n$ 维的矩阵。

根据这些定义，(19.3)中的目标函数能用方程表示为：

$$\pi = \underset{(1\times n)}{c'} \underset{(n\times 1)}{x}$$

这里，读者将会回想到：这个向量积 $c'x$ 是 $1\times 1$ 维的，因此，表示

的是一个标量。而涉及到 $m$ 个约束条件，矩阵符号的这种优势就会显而易见了，这是因为(19.3)中的约束条件的整个集合都可以缩写成以下简单的不等式形式：

$$\underset{(m\times n)}{A}\ \underset{(n\times 1)}{x}\ \leqslant\ \underset{(m\times 1)}{r}$$

这里，不等号应解释为相对应元素的不等，也就是，矩阵 $Ax$ 的第 $i$ 行小于或等于矩阵 $r$ 的第 $i$ 行，对于 $i$ 的每一个取值都适用。① 类似地，我们也可以将 $n$ 个非负限制用下面的一个单一不等式表示：

$$\underset{(n\times 1)}{x}\ \geqslant\ \underset{(n\times 1)}{0}$$

简言之，(19.3)中的线性规则可以简要地表示为如下形式： 664

$$\begin{aligned} &\text{Max. } \pi = c'x \\ (19.3')\qquad &\text{s.t. } Ax \leqslant r \\ &\text{且：} x \geqslant 0 \end{aligned}$$

根据同样的表达方式，很容易将(19.4)中最小化规划用如下方式表达出来：

$$\begin{aligned} &\text{Min. } C = c'x \\ (19.4')\ &\text{s.t. } Ax \geqslant r \\ &\text{且：} x \geqslant 0 \end{aligned}$$

## 关于求解的方法

在两个变量($n=2$)的情况下，用图解法求解会帮助我们很容易地得到最优解。我们确实不必考虑线性规划中的约束条件的个

---

① 不等号"≤"，当用于数字上时，通常与"≦"号是可以交换使用的。当用于向量时，这两种符号的意义就不同了。为便于讨论，可参阅 K.兰开斯特：《数理经济学》，纽约，麦克米伦出版社，1968 年，120 页。

数，这是因为附加的约束条件只能增加极点的个数，而并不增加图形的维数。而当存在三个选择变量时，这种方法就不太适用了，因为它需要有一个三维的图形。而对于维数更高的情况，这种方法将会彻底失灵！因此，我们必须找到一种能适合任何变量个数的非图解的求解方法。

首先，我们观察一下，两个变量情况下的求解程序是怎样逻辑地推广到 $n$ 个变量的情况的。对于 $n=2$ 的情形，我们的操作领域是两维空间(平面)。由于非负限制及约束条件的缘故，我们可以将研究领域变窄而只简单地将注意力集中在可行区域上，它是两维空间的一个子集。然后，通过目标函数，我们最后能在那个子集中确定作为最优解的特定点($\bar{x}_1$, $\bar{x}_2$)的位置。对于具有一般性的 $n$ 个变量的情形，我们必须代之以在 $n$ 维空间操作，其中每一点表示一个有序的 $n$ 元数组，($x_1$, $x_2$,…,$x_n$)或者是一个 $n$ 维向量。非负限制现在将我们限制在**非负的正交分划体**($n$ 维空间中的非负象限的类似物)之内，而那些约束条件将共同使这个作为可行区域的非负正交分划体线性化。最后，通过目标函数，我们可以在可行区域内确定这个作为最优解的特定点($\bar{x}_1$, $\bar{x}_2$, …, $\bar{x}_n$)的位置。

正如我们以前所熟知的那样，最优解总是可以在可行区域的极点处得到。正如下面将要说明的一样，即使是在 $n$ 个变量的情况下，这种结果也是成立的。因此，我们需要的不是要找出全部的可行域，而是要找到一种确定所有极点组成的集合的方法，从中我们就可以挑选出最优解。这种方便的结果是建立在这样的事实基础之上的：不必考虑选择变量的个数，线性规划中可行区域永远是一个闭合的凸集。

1　(19.3)中的 $m$ 个约束条件能写成如下形式吗？

$$\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} x_2 + \cdots + \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} x_n \leqslant \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}$$

2　(19.3)中的目标函数能否写成一个内积的形式？

3　假设(19.4)中的第一个约束条件恰好是一个等式，$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = r_1$，证明，通过用两个经过恰当选择的不等式代替这个等式，我们仍能在整个约束条件部分使用"$\geqslant$"号的格式。

## 19.3　凸集与线性规划

一个凸集，正如以前在 11.5 节中所定义的一样，是一个具有下述特点的点的集合：如果 $u$ 和 $v$ 是该集合中任意两点，那么 $u$ 和 $v$ 的任意凸组合也都在该集合中。用符号表示，如果

$$\left.\begin{array}{l} u \in S \\ v \in S \end{array}\right\} \Rightarrow \quad w \in S$$

其中　$w \equiv \theta u + (1-\theta) v \qquad (0 \leqslant \theta \leqslant 1)$

那么我们就称集合 $S$ 为凸集。这表明在一个线性规划的许多方面都有凸集出现。

以目标函数为例。对于取任意特定值的$\pi$(或 $C$)，都具有以下这些一般性的线性形式：

两维空间： $\pi_0 = c_1x_1 + c_2x_2$ [给出一条直线]

三维空间： $\pi_0 = c_1x_1 + c_2x_2 + c_3x_3$ [给出一个平面]

(19.6) $n$ 维空间： $\pi_0 = c_1x_1 + c_2x_2 + \cdots + c_nx_n$

[给出一个超平面]

直线和平面是超平面的特例，很容易看出，它们是凸集。不难证明，$n$ 维空间中的超平面也是一个凸集。

将满足(19.6)的点的集合称为集合 $H$，那么 $H$ 中的每个点都将位于由已知方程定义的超平面上。[1] 如果我们在集合 $H$ 中任选两点 $u$ 和 $v$，它们的坐标分别为($u_1$，$u_2$，…，$u_n$)和($v_1$，$v_2$，…，666 $v_n$)，那么，因为这些坐标一定满足方程(19.6)，所以，我们通过用 $u_i$ 代替 $x_i$，用 $v_i$ 代替 $x_i$，一定会发现下式成立：

$$\pi_0 = c_1u_1 + c_2u_2 + \cdots + c_nu_n$$

(19.7) $$\pi_0 = c_1v_1 + c_2v_2 + \cdots + c_nv_n$$

用更简洁的向量符号表示，可以写成：

(19.7′) $\pi_0 = c'u$ 和 $\pi_0 = c'v$

为了证明集合 $H$ 是凸集，有必要证明一下，$w$ 是 $u$ 和 $v$ 的任意凸组合，并且满足(19.6)，以便使 $\pi_0 = c'w$ 成立。所要求的证明相对简单，既然

$$c'w = c'[\theta u + (1-\theta)v] = c'\theta u + c'(1-\theta)v \quad [分配律]$$
$$= \theta c'u + (1-\theta)c'v \quad [向量乘法交换律]$$

---

① 而集合 $H$ 的正式定义要求的表达方式为：

$$H = \{(x_1, x_2, \cdots, x_n) \mid \pi_0 = c_1x_1 + \cdots + c_nx_n\}$$

或者，用向量符号表示为：

$$H = \{x \mid \pi_0 = c'x\} \quad [根据(19.5)]$$

集合 $H$ 可以通过(19.6)或者它的向量转换，表示成更简单的形式，$\pi_0 = c'x$。

$$= \theta\pi_0 + (1-\theta)\pi_0 \qquad [根据(19.7')]$$
$$= \pi_0$$

对于在闭区间[0,1]中的任意 $\theta$ 值,凸组合 $w$ 都在集合 $H$ 中。因此,集合 $H$ 的确是一个凸集。

如果考虑到超平面与其所处的 $n$ 维空间的关系,那么超平面总能将 $n$ 维空间分成两个半空间。例如,在两维的情况下,一条直线将坐标平面分为两个两维的半空间,它们分别位于直线的两侧。这一点,我们可以从图 19.1 和图 19.2 中所绘成的边界看出。类似地,一个 $n$ 维空间也可以被分成两个 $n$ 维半空间,但这需要一个超平面来分。我们是能得到一个开放的半空间,还是一个闭合的半空间,这要视是否将分割超平面看成是所研究的半空间的一部分而定。例如,如果我们写出如下两个不等式:

$$c'x < 9 \qquad 和 \qquad c'x \geqslant 9$$

那么,前者将定义一个位于超平面 $c'x = 9$ 一侧的开放的半空间,而后者将定义一个包括位于已知超平面另一侧以及超平面本身上的点的闭合的半空间。

根据以上讨论可知,(19.3)和(19.4)中的每一个约束条件都定义了一个闭合的半空间。进而,每个非负限制也是如此,因为不等式(如)$x_1 \geqslant 0$ 只不过是如下约束条件的一种特例而已:

$$(19.8) \qquad a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \geqslant r_1$$

令 $a_{11} = 1$ 而其它所有的系数(包括 $r_1$)均等于 0。结果,这些闭合的半空间中的每一个又都是一个凸集。

最后这句陈述的正确性在两维空间的情况下是相当显而易见的。在图 19.1 中,位于任意约束边界上或其右侧的点的集合是一个凸集,这是因为连接集合中任意两点的线段也一定在该集合中。

现在，让我们来说明，在 $n$ 维空间中的一个闭合半空间也是一个667凸集。考察一下，由不等式(19.8)定义的闭合的半空间，它也可以用向量符号表示为：

$$a'x \geqslant r_1 \text{这里 } a' \equiv [a_{11}a_{12}\cdots a_{1n}] \qquad (19.8')$$

令 $u$ 和 $v$ 是那个半空间中的任意两点。那么，因为 $u$ 和 $v$ 两个都满足(19.8′)，所以可以得出：

$a'u \geqslant r_1$ 和 $a'v \geqslant r_1$[参见(19.7′)]

对于 $0 \leqslant \theta \leqslant 1$，进而，我们可以推导出：

$$\theta a'u \geqslant \theta r_1 \text{和 } (1-\theta)a'v \geqslant (1-\theta)r_1 \qquad (19.9)$$

现在，令 $w = \theta u + (1-\theta)v$ 是 $u$ 和 $v$ 的凸组合。如果可以证明 $w$ 也满足(19.8′)，那么所研究的这个半空间一定是凸的。为此，我们可写成如下向量积：

$$a'w = a'[\theta u + (1-\theta)v] = \theta a'u + (1-\theta)a'v$$

但，根据(19.9)，很容易得出：

$$a'w \geqslant \theta r_1 + (1-\theta)r_1 \quad \text{或} \quad a'w \geqslant r_1$$

因此，像 $u$ 和 $v$ 一样，凸组合 $w$ 也满足(19.8′)。这证明：$n$ 维空间中的闭合半空间是(闭合的)凸集。

将这条推理更深入一些，我们也能得出这样的结果：一个具有一般性的 $n$ 个变量的线性规划的可行区域是一个闭合的凸集。作为一个初步的结论，我们注意到，可行区域总可表示总共 $m+n$ 个闭合凸集的交集。一般而言，根据定义，可行区域中的任一点一定同时满足 $m+n$ 个线性(弱的)不等式组——$m$ 个约束条件加上 $n$ 个非负限制。因此，它一定同时也是 $m+n$ 个闭合半空间中的一员，也就是说，它一定是那些 $m+n$ 个闭合凸集的交集中的一点。鉴于这种情况，下述定理将确立可行区域为一个闭合凸集：

有限个凸集的交集是一个凸集,且如果其中每一个集合都是闭合的,那么其交集也将是闭合的。

这个定理的本质性内容可以从图 19.4 中领会到,这里集合 $S$(一个正方形)和集合 $T$(一个正三角形)两个都是凸集。它们的交集 $S \cap T$,这里用最暗的阴影区域表示,很显然也是凸集。再 668 者,如果 $S$ 和 $T$ 两个都是闭合的,那么 $S \cap T$ 也将是闭合的,这是因为交集的边界点,它们仅仅是集合 $S$ 和集合 $T$ 边界点的一个子集,确实属于交集。[1]

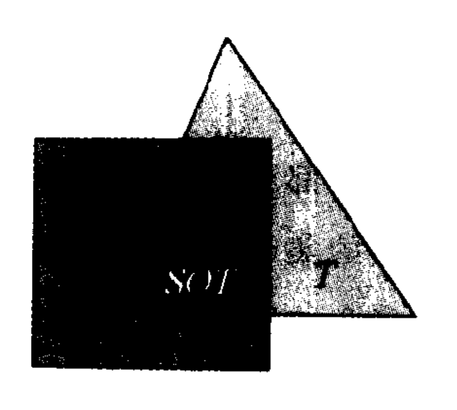



图 19.4

## 极点与最优解

从前面的论述中,我们可以得出两个主要结论:(1)对于给定任一值的$\pi$(或 $C$),一个 $n$ 变量线性规划的目标函数总能定义一个超平面,且这个超平面是一个闭合的凸集。(2)可行区域,由于其为 $m+n$ 个闭合半空间的交集,所以也是一个闭合的凸集,称该集合为 $F$。现在,这两个结论是彼此相联系的。

在进一步探讨之前,我们有必要先区别一下内点,边界点和极

---

① 然而,观察一下,两个凸集的并集却不一定是凸集。在图 19.4 中,其并集 $S \cup T$ 由整个阴影区域组成,它很明显是凹集。

点的概念。边界点和内点的区分是显而易见的。参见图 19.4,位于正方形四个边上的点是集合 $S$ 的边界点,而那些没有位于边上的点是内点。用更正规的形式表述,即任何集合 $S$ 中一个边界点 $b$ 必定具有这样的性质:$b$ 的每一个邻域,不论它有多么小,一定存在不属于集合 $S$ 的一个或多个点。另一方面,集合 $S$ 的一个内点 $i$ 具有这样的性质:点 $i$ 的足够小的邻域将包含那些只属于集合 $S$ 的点。[①] 至于极点,它们只不过是边界点的一种特例而已。之所以特殊,是因为极点是一种并不在该集合中任意两点所连成的线段上的点。也就是说,极点并非是来源于该集合中任意两点的凸组合的一种点。再次参见图 19.4,正方形的四个角点符合极点的定义,而其它所有的点都不符合极点的定义。简言之,给定任一集合 $S$,$S$ 的所有极点的集合是 $S$ 的所有边界点集合的一个子集。但是,所有边界点集合与所有内点集合并不相交。

在力图达到最优化的过程中,我们的目标是在集合 $F$ 中通过改变$\pi$或 $C$ 值,将目标超平面或者“推到”可能的最高位置(达到 $\bar{\pi}$),或者“推到”可能的最低位置(达到 $\bar{C}$)。当达到最佳位置时,该超平面 $\bar{H}$ 可以不包含集合 $F$ 的内点;这是因为,如果它包括集合 $F$ 的内点,那么我们总能将其“推”得更远些,从而达到一个更好的位置。因此,只有集合 $F$ 的边界点才能出现在集合 $\bar{H}\cap F$ 中。这就引出了支撑超平面的概念。

一个支撑超平面,(比如 $\bar{H}$)是一个与凸集($F$)有一个或多个公共点的超平面,而它所处的位置,使得集合 $F$ 只位于 $\bar{H}$ 的一侧。

---

① 这里在应用邻域的概念方面,我们应记住问题中我们所定义的那些集合所在空间的维数。例如,在图 19.4 中,因为这些集合是在两维空间中定义的,所以我们所提到的任何一个邻域本质上都应是两维的。

图 19.5 用两维和三维的例子说明了这个概念。在图 $a$ 中,直线 1、2 和 3 是支撑超平面(这里是直线)的例子。直线 1(或者直线 669 2)与正多边形只有一个公共点,而直线 3 却有 $n$ 个公共点。但,无论是哪种情况,集合 $F$ 都只位于支撑直线的一侧。因此,只有 $F$ 的边界点能出现在这些直线中的每一条上。注意,每条支撑直线至少包含集合 $F$ 的一个极点,例如 $f,g$ 和 $h$。图 $b$ 中的三维示例也是相似的,只是现在将支撑直线换成了支撑平面。再者,$\bar{H}$ 和 $F$(这次是正多边形)的交集只能由 $F$ 的边界点构成。正如图所示,仅仅包含一个点($u$),而此点是集合 $F$ 的一个极点。

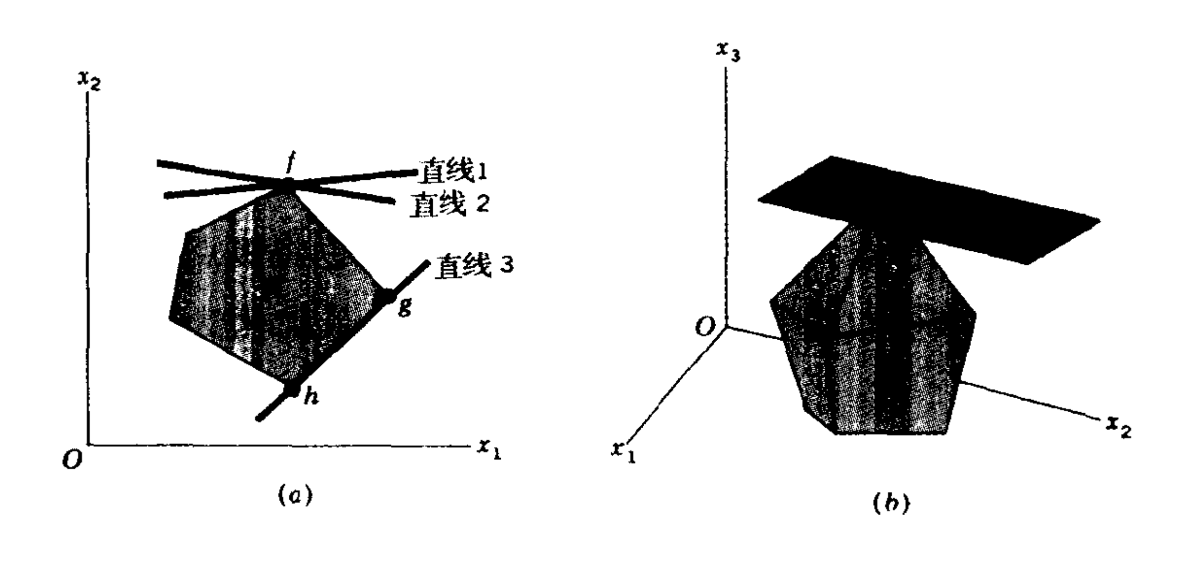



图 19.5

对于具有一般性的 $n$ 维空间的情况,我们关于图 19.5 讨论的本质内容可以表述成下述两个定理:

**定理 I** 给定一个闭凸集的一个边界点 $u$,在 $u$ 处至少存在着一个支撑超平面。

**定理 II** 对于一个下有界的闭凸集,在每个支撑平面中,至少存在一个极点。

这两个定理与线性规划的关系是显而易见的。当我们找到最

优解时，代表最优的等利润线或等成本线的那个目标超平面将是一个支撑超平面。根据定理 I，可行区域中的每个边界点都有可能是最优解。但是，定理 II 通过使我们只注意到**极点**，从而缩小了问题的范围。即使在同一个支撑超平面中，还存在不是极点的边界点，但是，这些点也总是与有相同$\pi$（或 $C$）值的极点相联系，因此，它们也不会比极点更好，所以，不考虑它们也不会造成任何损失。

这个简单但却重要的事实应用在解决 $n$ 变量的线性规划的**单纯形法**中是有很大益处的。它最初是由乔治·B.丹蒂格[①] 提出来的，对此我们将在下节中解释。所谓**单纯形**是在 $n$ 维空间中，一种三角形的类似物，具有代表极点的隅角。这种单形法给我们670提供了一个系统化的方法，由此，我们可以从可行区域中的一个极点移到另一个极点，直到找到最优解。我们可以将极点的选择区域缩小这一事实仅仅是线性规划区别于古典最优化法的特点之一。

### 局部最优与总体最优

线性规划区别于古典最优化问题的另一个重要性质是其所得到的任何一个解给出的不仅是局部的（相对的），而且也是总体的（绝对的）最优。下述定理（整体性定理）将会对产生这一特征的原因作出解释。该定理虽然并未给出在下述条件下，局部最优也可以作为全局最优的必要条件，但却给出了充分条件：

---

① 乔治·B.丹蒂格："满足线性不等式的变量线性函数的极大化"，收载于 T.C.库普曼主编的《生产与配置的活动分析》，John Wiley & Sons. Inc.，纽约，1951 年，339—347 页。

如果可行集合 $F$ 是一个闭凸集，并且目标函数在可行集合中是一个连续的凹(凸)的函数，则：($a$)任何一个局部极大值(极小值)也将是整体极大值(极小值)，且($b$)集合 $F$ 中使得目标函数最优化的点将构成一个凸集。如果目标函数在可行集合中恰好是严格凹(凸)的，那么整体极大值(极小值)将是唯一的。

这个定理的本质性内容，我们以前在图 11.5 中已经遇到过。它与现在这种情况的特殊关系在于：这些关于整体最优化的充分条件在线性规划中当然是很适用的，但是，在古典最优化问题中，它就不成立了。正如我们所观察到的，线性规划中的可行区域总是一个闭凸集。而且，用选择变量表示的目标函数是连续的和线性的。它既可以被看作是凹函数，又可以被看作是凸函数，这取决于问题中所要求的是最大化问题还是最小化问题。因此，定理中所陈述的充分条件确实能得到满足，并且所得的任何一个最优解本质上都是总体的。这与定理中的($a$)部分是一致的。

为了理解该定理的($b$)部分，让我们再看一看图 19.5。如果线性规划的最优解是在某一单独的极点处得到的，例如，图 $a$ 中的 $f$ 点或图 $b$ 中的 $u$ 点，那么，依照惯例，那个点本身就构成一个凸集，而且该定理中的这种断言很快就能得到证实。但是，如果有多重最优存在时，该如何呢？多重最优是在当支撑超平面与可行集合的接触点不只是一个点时，才出现的。例如，当直线 3(图 $a$)是一条支撑直线，或者当支撑平面 $\bar{H}$(图 $b$)与正多面体 $F$ 的一个面(平面)重合。在这些例子中，所有最优化点所组成的集合既可以转化在线段 $hg$ 上，又可以转化在正多面体 $F$ 的某一个平面上，其他情况也是如此。因此，正如定理所断言的，该集合又是一个凸

671

集。鉴于此，我们可以确信，如果对于一个线性规划存在一对最优解（两个解同样最优），那么这两个解的任何凸组合或加权的平均值也一定是一个最优解。

## 19.4 单纯形法：求极点

一个线性规划的最优解可以在极点中得到。如果给定一个可用图示的可行区域 $F$，那么要找到极点是很容易的。但是，对于那些无法图示的 $n$ 个变量的情形又该如何呢？让我们再回过头来考察一下具有两个变量情况的极点，以便能尽可能找到一些线索。

### 松弛与剩余

图 19.1 和 19.2 中的极点分属于三种主要类型。这些类型可在图 19.6 中得到充分地说明。该图重新构造出图 19.2*b* 的可行区域。

第一种类型的极点是由两个约束边界的交点构成的。这些例子可在点(8,8)和点(16,4)处找到。并且这样的点确实能完全地满足那两个约束条件，但它却不一定能完全地满足余下的约束条件。例如，点(16,4)能完全地满足切割和包装的约束条件，但却不能完全地满足混合的约束条件，这是因为该点位于混合边界的下方。由于一些约束条件得不到完全满足，所以一定存在着某种能力未得到充分利用的情况。（或者，在饮食问题中，过度摄取一些营养物质而超过最低需求量。）也就是说，在能力的利用方面的松弛（或在营养物质摄取方面产生剩余）有待于改进。

第二类极点，以(0,8)和(16,0)为例说明。它们产生于一条约

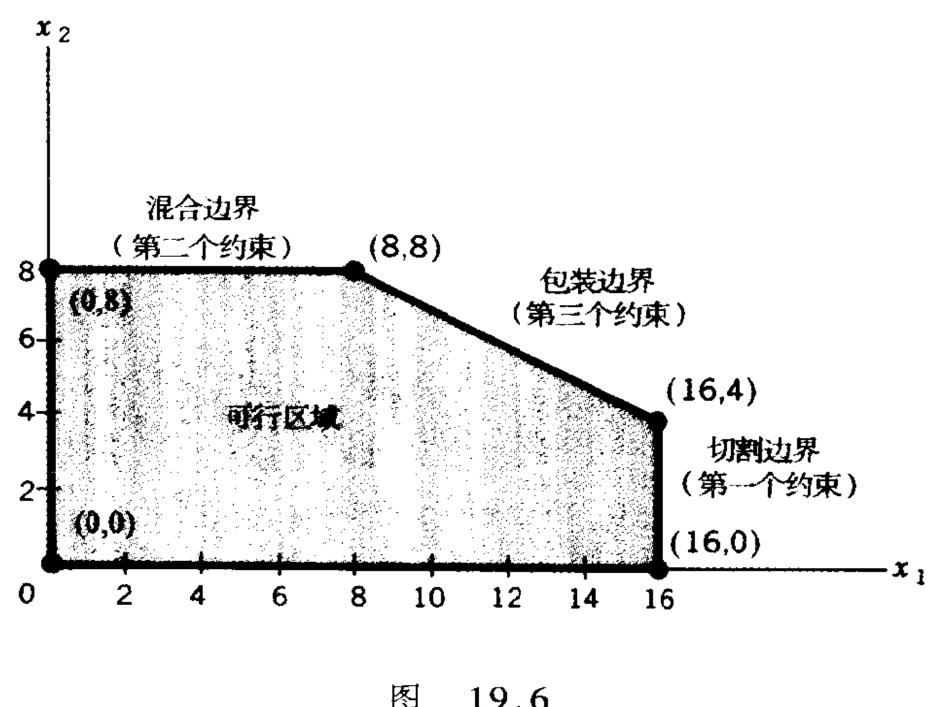



图 19.6

束边界与一个坐标轴的交点处。因为它只位于一条约束边界上，所以这样的点仅仅能完全满足一个约束条件。鉴于这种差异，现在就会在两个约束条件中产生松弛。

最后，作为第三类极点，我们得到的是原点(0,0)，它对每个约束条件都不能完全满足。然而，极点(0,0)只有在最大化规划中存在，因为最小化规划中的可行区域一般均不包括原点，这一点读者可以从图 19.1*b* 中看出。

结果是，在现在这个例子中，这里的约束条件的个数（3）超过了选择变量的个数（2），至少在其中一个约束条件中，每个极点都包含着一个松弛。而且，这一点在图19.6 中是很明显的，由每个极点所引起的松弛的特定数量是很容易计算出的。当我们确定某一点为最优解时，我们不仅确定了 $\bar{x}_1$ 和 $\bar{x}_2$ 的值，而且还确定了松弛的最优值。让我们力图简明地考察一下松弛，并且用符号 $s_i$ 表示第 $i$ 个约束条件中的松弛。虽然在最小化问题中 $s_i$

会变成剩余变量,但是 $s_i$ 仍代表松弛变量。这些变量都被称作**虚拟变量**。

简单地考察一下松弛和剩余,将使我们可以将约束条件的每个不等式都转换成一个严格的方程。更重要的是,它将能引导我们找出一种代数方法来求出可行区域中的极点。

### 经过转换的线性规划

让我们再返回到(19.2)中的生产问题,它的可行区域已经在图 19.6 中描述过。我们可以通过向每个约束条件上附加上一个松弛变量并且适当地改变一下目标函数和非负限制,我们就可以将那个线性规划改写成以下形式:

$$
\begin{aligned}
&\text{Max. } \pi = 40x_1 + 30x_2 + 0s_1 + 0s_2 + 0s_3 \\
&\text{s.t. } x_1 + s_1 = 16 \qquad [\text{切割}] \\
&x_2 + s_2 = 8 \qquad [\text{混合}] \\
&x_1 + 2x_2 + s_3 = 24 \qquad [\text{包装}] \\
&\text{并且: } x_1, x_2, x_3, s_1, s_2, s_3 \geqslant 0
\end{aligned}
\tag{19.10}
$$

673 读者注意,(19.10)中这三个经过转换的约束条件还可以用如下的矩阵方程表示:

$$
\begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 2 & 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} x_1 \\ x_2 \\ s_1 \\ s_2 \\ s_3 \end{bmatrix}
=
\begin{bmatrix} 16 \\ 8 \\ 24 \end{bmatrix}
\tag{19.11}
$$

现在一共有5个变量。松弛变量 $s_i$ 与 $x_1$ 和 $x_2$ 一样被限定为非负。当 $s_i>0$ 时,在第 $i$ 个约束条件中存在着松弛。当 $s_i=0$ 时,第 $i$ 个约束条件得到完全满足。但 $s_i$ 永远非负。在目标函数中 $s_i$ 之所以被赋予零系数,这是因为松弛对利润无贡献。事实上,它们可以被一起从目标函数中删掉。还应注意,在最小化的问题中,这个(非负的)虚似变量一定会代之以 $-s_i$ 的形式出现在约束条件中。

我们很容易求出由每一极点所隐含的松弛值。例如,对于点(0,0),我们可以将 $x_1=0$ 和 $x_2=0$ 代入到这三个经过转换的约束条件中,得出:$s_1=16$, $s_2=8$ 和 $s_3=24$。因此,图 19.6 中的两维产量空间中的点 $(x_1, x_2)=(0,0)$,映射到五维解空间中点,

$$(x_1, x_2, s_1, s_2, s_3) = (0,0, 16, 8, 24)$$

如表 19.3 中第一行所示。依照同样的程序,读者也可以验证表中所列出的另外四个极点的映射。

关于表 19.3 的结果,值得注意的是这样的事实:在该解空间中表示极点的所有五个点都具有这样一种共性,即这五个变量中恰好有三个具有非零值。当然,这不仅仅是巧合。而且,“三”不是其它任何数目,而是我们线性规划中约束条件的精确个数。

为了便于理解这个结果,让我们回忆一下以前关于方程和变量个数的计算方面的讨论。通常,假定 $m$ 个方程的方程组相容且无关,我们可以预期,只有当其正好存在 $m$ 个变量时,我们才能得出确定的解。在我们现在这种情况下,由于在(19.11)中有三个应考虑的约束条件,也就是 $m=3$。因此,在五个变量中,具有确定

674 的非零解的变量个数不应超过三个。[①] 当我们考察以下约束条件的向量方程转换时，这一点也许会更清楚些。

表 19.3

| 产出空间 |  | 解空间 |
|---|---|---|
| $(x_1, x_2)$ | → | $(x_1, x_2, s_1, s_2, s_3)$ |
| (0,0) |  | (0,0,16,8,24) |
| (16,0) |  | (16,0,0,8,8) |
| (16,4) |  | (16,4,0,4,0) |
| (8,8) |  | (8,8,8,0,0) |
| (0,8) |  | (0,8,16,0,8) |

$$\begin{bmatrix}1\\0\\1\end{bmatrix}x_1+\begin{bmatrix}0\\1\\2\end{bmatrix}x_2+\begin{bmatrix}1\\0\\0\end{bmatrix}s_1+\begin{bmatrix}0\\1\\0\end{bmatrix}s_2+\begin{bmatrix}0\\0\\1\end{bmatrix}s_3=\begin{bmatrix}16\\8\\24\end{bmatrix}\qquad(19.11')$$

如果我们令这五个变量中的任意两个变量为零，那么，就相当于从(19.11′)的左边删掉两项，其结果是将得到一个由三个变量表示的三个方程联立的方程组。倘若左侧那些（保留下的）系数向量是线性无关的，那么就可以得到一个唯一解。当这三个变量的解值与另两个被预先赋予零值的变量放在一起时，就会出现与表 19.3 所示一样的一个 5 元有序数组。

## 基本可行解和极点

然而，刚刚描述的步骤，可产生两种可能的情况。首先，也许会出现负的解值。例如，如果我们令 $x_1 = s_3 = 0$，那么(19.11′)将

① 但在解空间中，也可能会遇到取非零值的变量数少于 $m$ 个的情况。此即所谓**退化**现象，我们将在 19.6 节对其进行扼要讨论。

得出：$x_2=12$，$s_1=16$ 和 $s_2=-4$。因为这违背了非负限制，所以这样的解显然是不可行的，一定要舍去。用图表示，整个特定的不可行解表示的是包装边界与垂直轴的交点，它显然位于可行区域以外，是位于集合 $F$ 的补集 $\overline{F}$ 中。

其次，所有的解值也许都是非负的。在这种情况下，解将与可行区域极点中的一个相一致，并且该解将构成线性规划的一个基本可行解(BFS)。该解之所以是可行的，是因为它位于可行区域内：它既满足约束条件又满足非负限制。之所以又称其为基本的是因为该解的获得是依三个线性无关的系数向量的存在而定的，而这三个向量一起构成了三维空间的一个基底。(这个三维空间，被称作需求空间，与约束条件有关；当然，它的维数是由我们所要求的约束条件的个数决定的。)由于上述的对应性，因此现在搜寻极点就相当于搜寻经过转换的约束方程的基本可行解。

举一例子加以说明。令(19.11′)中的 $x_1=x_2=0$，则

$$\begin{bmatrix}1\\0\\0\end{bmatrix}s_1+\begin{bmatrix}0\\1\\0\end{bmatrix}s_2+\begin{bmatrix}0\\0\\1\end{bmatrix}s_3=\begin{bmatrix}16\\8\\24\end{bmatrix} \tag{19.12}$$

或等价于：

$$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}s_1\\s_2\\s_3\end{bmatrix}=\begin{bmatrix}16\\8\\24\end{bmatrix} \tag{19.12′}$$

675

因为，最左侧的矩阵是一个单位矩阵，去掉它并不影响方程的解，所以就可以直接将解读出：$s_1=16$，$s_2=8$ 和 $s_3=24$。这个唯一解的存在是由三个系数(列)向量的线性无关性作保证的，而这三个系数向量是生成一个三维空间(需求空间)的单位向量。由于

解是非负的，所以它是一个 BFS。并且，因为该解值在解空间中形成了点(0,0,16,8,24)或是图 19.6 中的点(0,0)，所以 BFS 的确与某个极点是一致的。

令(19.11′)中另外一对变量为零很明显会产生一个全新的方程组。如果该系数向量又是线性无关的，那么我们就可以在其需求空间中得到一个新的基底，而且若解是非负的，还可以得到一组新的 BFS。这又为我们确定了一个极点。因此，在可行区域中，向不同极点的移动实质上意味着能够在三维的需求空间中快速地找到一组新的基底。这是我们后面讨论中要利用的一个基本事实。

有关 BFS 这个概念，重要的是，确定 BFS 的方法是一种代数的方法而不是几何的方法，这样便可以易于推广应用到具有 $m$ 个约束条件，$n$ 个选择变量的一般性的线性规划中。在后面这种情况下，需求空间将是 $m$ 维的，而解空间将是 $m+n$ 维的。为了找到一个 BFS，我们现在在经过转换的约束方程中需令 $m+n$ 个变量中的 $n$ 个为零。但其方法与(19.12)中所说明的那种较简单的情况没有什么不同。

## 练习 19.4

---

1 将下列线性规划转换成(19.10)的形式：

(a) Max. $\pi = 3x_1 + 2x_2 + 5x_3$

s.t.

$$
\begin{aligned}
3x_1 + x_2 \quad\quad &\leqslant 10 \\
x_2 + 2x_3 &\leqslant 6 \\
2x_1 + x_2 + x_3 &\leqslant 8
\end{aligned}
$$

并且 $x_1, x_2, x_3 \geqslant 0$

(b) Min. $C = x_1 + 6x_2 + 2x_3$

s.t. $x_1 + 2x_2 \geqslant 2$

$x_1 + x_2 + 3x_3 \geqslant 12$

并且 $x_1, x_2, x_3 \geqslant 0$

2 利用虚拟变量转换练习 19.1-1 中的线性规划

3 上题中的线性规划分别具有如下两组极点：

(a) (0,0),(0,3),(2,3),(4,2),(4,0)

(b) $(4,0),(2,\frac{1}{2}),(\frac{3}{5},\frac{6}{5}),(0,3)$ 676

通过将这些($x_1$, $x_2$)值代入适当的经过转换的约束条件中，求出对应于每个极点的 $s_i$ 的值。然后，如表 19.3 中一样，将其表示成五维解空间中的点。

4 上题中的极点用有序 5 元数组的形式表示时，你发现每个点中有几个非零元素？这种结果是巧合吗？

5 假定有一个圆盘(密实的圆)，令 $I$ 为其所有内点的集合，$B$ 为其所有边界点的集合，$E$ 为其所有极点的集合，则下述哪些说法是正确的？

(a) $I \subset E$ (d) $B = E$

(b) $B \subset E$ (e) $I \cap B = \phi$

(c) $E \subset B$ (f) $E \cup I = B$

6 给定一个具有下列条件的线性规划，求其要求空间与解空间的维数各是多少：

(a) 7 个选择变量和 3 个约束条件？

(b) 3 个选择变量和 5 个约束条件？

(c) $n$ 个选择变量和 $m$ 个约束条件？

---

## 19.5 单纯形法：求最优极点

给某一极点定位就是要找到某个基本可行解。但我们怎样才能找到最优极点和 BFS 呢？在低维数的规划中，我们当然可以找出所有的 BFS 的规划，计算出相应的被求极大值函数(或被求极

小值函数)的值,然后从中挑选出最优的一个。当涉及到许多变量及约束条件时,这种方法需要的计算量很大。单纯形法的概念就是从某一初始点出发,计算出目标函数的值,然后通过向邻近极点运动来观察后者(目标函数值)是否有改进。如果能得到改进,我们就可以进行这次移动。然后,通过进一步的移动来决定是否会有进一步的改进。当我们最终达到一个不再会有进一步改进的某一极点时,它将构成最优解。这种从可行区域的一个隅角运动到另一个隅角的迭代过程之所以可以减少计算量,是因为在该过程中只需考虑所有极点所组合的集合的一个子集。

我们将通过进一步讨论(19.10)中的那个经过转换的最大化问题的线性规划,来说明这种方法。

## 单纯形表

677 在(19.12)和(19.12′)中,我们已经通过令 $x_1=x_2=0$ 找到了一个原始的 BFS。在解空间中就产生了如下有序 5 元数组,$s_1$ 及与其相对应的利润数$\pi_1$:

$$s_1=(0,0,16,8,24)$$
$$\pi_1=40(0)+30(0)=0 \tag{19.13}$$

**表 19.4 单纯形表:表 I**

| | $\pi$ | $x_1$ | $x_2$ | $s_1$ | $s_2$ | $s_3$ | 常数 |
|---|---|---|---|---|---|---|---|
| 0 行 | 1 | -40 | -30 | 0 | 0 | 0 | [0] |
| 1 行 | 0 | ① | 0 | 1 | 0 | 0 | 16 |
| 2 行 | 0 | 0 | 1 | 0 | 1 | 0 | 8 |
| 3 行 | 0 | 1 | 2 | 0 | 0 | 1 | 24 |
| | | | | * | * | * | |

利用如表19.4中所示的所谓的单纯形表,按照图示的方法也能得到同样的答案。①

在这个表中,我们将每一个选择变量及松弛变量的专门给出一列。此外,还设有一个$\pi$值的列和一个常数列。包含$\pi$列可使我们既能够具体地了解到表中包含在目标函数中(0行)的信息,又能了解到三个经过转换的约束条件(1行,2行,3行)的数据。每个变量下的所有数字仅仅是相关方程中那个变量的系数,而常数列中的数字不属于任何一个变量。常数列左侧的那条垂直直线在各个方程中应是"="号的位置。因此,0行就应读作$\pi-40x_1-30x_2=0$,这是目标函数的一种简单变形。类似地,1行可写成:$x_1+s_1=16$(第一个约束条件)等等。它也许还有助于发现,如果不考虑0行和$\pi$列,表中剩下的数字恰好是经过适当转换的(19.11)的那些数字。

根据我们以前的讨论,找到一个BFS就是要找到对于三维需求空间的一个特定的基本解。这意味着,如果我们暂时不考虑0行,那么我们必须从后三行中找到三个线性无关的列向量。很显然应选择带星号的列,它们一起构成了一个3×3维的单位矩阵。因此,我们可以将$s_1,s_2$和$s_3$考虑进基本解中,因而令$x_1=x_2=0$。如果将$x_1$和$x_2$列有目的地从此表中删掉,那么下部的三行恰好简化成(19.12′)的形式,而得到,$(s_1,s_2,s_3)=(16,8,24)$。它是非负的,因此是一个BFS。

由于产量为零,所以利润也是零,很显然,这个BFS不是最优

① 以下所描述的是修正的单纯形法,它不同于单纯形法之处在于,表中包含了目标函数。

的。但在最大化问题中,能够采用只包括松弛变量的解作为初始基本可行解总是非常理想的。松弛变量总有线性无关的单位向量作为其系数向量,因此它们自然能提供一个现行的基。

678

事实上,利润信息也可以从表中直接看到。因为若将0行与下面的3行一并考虑(但仍不考虑 $x_1$ 和 $x_2$ 列),则该表变成方程组

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \pi \\ s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 16 \\ 8 \\ 24 \end{bmatrix} \tag{19.14}$$

它不仅揭示出松弛变量的值,而且也给出额外信息 $\pi_1 = 0$(这里下标1是指初始BFS。一般而言,当表中出现 $4\times4$ 单位矩阵时,如表19.4,就可以确定一个BFS,而且与该BFS相关的目标函数的值,以及需求空间基变量的值,均可一并从表中读出。比较(19.14)与表19.4可知,表中常数列最上面的那个元素(圈起来以示强调)就是所求的利润值。为了解利润信息,今后,我们总是考察 $4\times4$ 矩阵而非 $3\times3$ 单位矩阵;或者更一般地,应考察 $(m+1)\times(m+1)$ 矩阵,而非 $m\times m$ 矩阵。

## 一个旋转步骤(a pivot step)

现在,我们通过形成一个新的基,转换到新的BFS,以争取提高利润。变换基的过程通称为旋转,其核心思想是由现在的不包括在基中的一个列向量取代基中某个列向量。或者是说,我们必须放弃目前包含的一个变量($s_1$,$s_2$,或 $s_3$),而采用现在未包含的

一个变量($x_1$ 或 $x_2$)。那么,我们是依据什么标准选择换出和换入向量(或变量)呢?

因为目的是要提高利润,所以,我们自然要向目标函数寻求线索。由(19.10)可知,目标函数表明 $x_1$ 的边际利润是 40 美元,$x_2$ 的边际利润是 30 美元。自然,选择 $x_1$ 作为换入变量,增加利润的希望更大。按照表 19.4,我们的标准就是应找出这样的变量,它在 0 行中有绝对值最大的负的表值。这里合适的表值是 $-40$,进入变量是 $x_1$;我们把 $x_1$ 列称为主元列。

主元列肯定会替换 $s_i$ 列中某一列。但这又产生两个问题。第一,我们必须确定 $s_i$ 列中哪一列退出。第二,作为新的基,主元列必须与保留下来的原向量线性无关。把主元列变为单位向量,该列下面 3 行中的元素,一个为 1,其余为零时,恰巧这两个问题 679 都可以解决。我们暂时不考虑变换的方法,而先考察变换的含义。如果变换后的主元列在第 1 行的元素为 1,因而它看起来正好像 $s_1$ 列。当然,令 $s_1$ 为换出变量,因为这种替换能保持基线性无关。同样,如果单位元素在第 2 行,那么我们令变换后的主元列替换 $s_2$ 列。两个关系密切的问题,即选择换出变量和保持线性无关的问题,用上述办法总可以立即得到解决。

在主元列中,单位元素的确切位置仍有待于确定。为方便起见,主元列中等于 1 的元素称为主元(pivot element)。选择主元元素主要考虑的是必须保持在三个车间的生产能力限度以内。例如,如果让第 3 行中的元素为主元,那么在变换为单位向量以后,$x_1$ 列将与现在的 $s_3$ 列相同。这意味着 $x_1$ 取代第二个基可行解中的变量 $s_3$,正如我们从(19.14)中所看到的,$x_1$ 的解值为 $x_1=24$(替换 $s_3=24$)。但根据(19.10),这个产出水平违背切割约束,

因此是不可行的。用另一种方法也可看出这一点：如果 $x_1$ 替换 $s_3$，新表中将在常数列的第 1 行（切割约束）中出现负元素，因而所得到的解是不可行的，必须排除。另一方面，如果我们选第 1 行的元素（标有圆圈）作为主元，那么变量 $x_1$ 取代变量 $s_1$，$x_1$ 的解值为 $x_1=16$（替换 $s_1=16$）。因为这个（较少的）产出的确不违背（19.10）中任何约束条件，所以确实可以把标有圆圈的元素（1）作为主元。在这种情况下，新表中常数列的所有常数保持非负。注意，在这方面我们不必担心 2 行中的 0 元素；零系数表示变量 $x_1$ 与第二个约束条件均无关，因而变量 $x_1$ 用作新基决不会遇到来自这个特殊约束条件的任何阻碍。

如果这样选取主元，就能很好地照顾到上面所提到的约束条件的对约束条件的考虑：（1）在主元列中除 0 行以外挑选出都为**正**的各行元素。（2）用每个正元素除常数列中相应的元素。（3）我们把所得的商称为**位移商**（displace*ment quot*i*ents*）。比较这些商，以其最小商所在的行为**主元行**。（4）把主元和主元列交点的元素选作主元。

通过选出最小位移商的方法，就能保证增加的产出（$x_1$）足够小，从而保持在所有约束范围之内，且能够确保所有变量都不会违背非负限制。在本例中，只有两个位移商可供比较：16/1 和 24/1。因为前者最小，所以 1 行应是主元行，其中围有圆圈的元素（1）为主元。

680 下一步，就是把主元列变成单位向量，使其主元等于 1，其他元素等于零。在本例中，主元已等于 1，因而不需再做什么了。一般而言，如果主元是数 $k$，可用 $k$ 遍除主元行中所有元素，包括常数列的元素，把它变为 1。（这相当于用同一常数遍除某个约束方

程的所有各项。)2 行元素已是零,也不再需要运算。为了使 0 行中的元素 - 40 变成零,把主元行(1 行)的 40 倍加到 0 行。(这就相当于把一方程的常数倍加到另一方程,决不影响方程相等。)同样,为了把 3 行中的元素 1 变成零,可将 3 行减去主元行(1 行)。在表 19.5 中给出这些结果,这样我们得到一个新单纯形表(表 II)。

如果读者想要去掉 $x_2$ 和 $s_1$ 列,可将表 II 视为表示下列方程组:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \pi \\ x_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} 640 \\ 16 \\ 8 \\ 8 \end{bmatrix} \tag{19.15}$$

**表 19.5 单纯形表:表 II**

| | $\pi$ | $x_1$ | $x_2$ | $s_1$ | $s_2$ | $s_3$ | 常数 |
|---|---|---|---|---|---|---|---|
| 0 行 | 1 | 0 | -30 | 40 | 0 | 0 | 640 |
| 1 行 | 0 | 1 | 0 | 1 | 0 | 0 | 16 |
| 2 行 | 0 | 0 | 1 | 0 | 1 | 0 | 8 |
| 3 行 | 0 | 0 | ② | -1 | 0 | 1 | 8 |

由这个方程,容易读出四个变量的解值,因为左边的单位矩阵可以去掉。当然,单位矩阵的出现决不是巧合,因为得到方程(19.15)解的变换过程,实质上就是用矩阵求逆的方法求解方程组的过程。但这一次不求出逆矩阵,我们已直接得到单位矩阵 $AA^{-1} = I$。

实际上,不写出(19.15),也可直接从表 II 的常数列中读出解值。在$\pi$列中的单位向量的单位元素是在 0 行;于是 0 行中的常数也可看作$\pi$的解值。同样,也可读出新基中变量 $x_1$,$s_2$ 和 $s_3$ 的解 681

值。另一个方面，$x_2$ 和 $s_1$ 列不含单位向量，没有进入基中，其解值自然为零。因此，第二个基可行解必为：

$$S_2 = (16,0,0,8,8)$$

$$\pi_2 = 640 \tag{19.16}$$

与 $S_1$ 比较时，现在我们得到的利润大得多(640 与 0 对比)，根据表 19.3 进行这个旋转步骤，我们已由第一个极点转到第二个极点。

## 另一个旋转步骤

在表 II 的 0 行中有一个表值 -30，与变量 $x_2$ 有关。因 0 行中负表值表示正边际利润率，让 $x_2$ 替换基中边际利润率为 0 或为负的变量，可进一步提高利润。因此，我们选 $x_2$ 列作为下一个主元列。至于主元行，因为最小的位移商是

$$\min\left\{\frac{8}{1},\frac{8}{2}\right\} = \min\{8,4\} = 4$$

所以应选 3 行为主元行。因此主元是 2，我们已用圆圈围住以引起注意。

为了把 $x_2$ 列中的元素变成 0,0,0 和 1(按这个次序)，我们必须：(1)把 3 行(主元行)的 15 倍加到 0 行；(2)1 行保持不变；(3)2 行减去 3 行的 1/2 倍；(4)3 行除以 2。希望读者一定完成这些运算并对照表 19.6 中的表 III 校验其结果。

表 III 中单位向量所在的列与变量 $\pi$，$x_1$，$x_2$ 和 $s_2$ 有关，其解值读出如下：$\pi = 760$，$x_1 = 16$，$x_2 = 4$(由 3 行读出)和 $s_2 = 4$(从 2 行读出)。因此，可写成

$$S_3 = (16,4,0,4,0)$$

$$\pi_3 = 760 \tag{19.17}$$

现在生产的两种产品,其利润可提高到760。另外,由表19.3看出,第二个旋转步骤把我们带到了可行区域的新极点。

**表19.6 单纯形表:表III**

| | $\pi$ | $x_1$ | $x_2$ | $s_1$ | $s_2$ | $s_3$ | 常数 |
|---|---|---|---|---|---|---|---|
| 0行 | 1 | 0 | 0 | 25 | 0 | 15 | 760 |
| 1行 | 0 | 1 | 0 | 1 | 0 | 0 | 16 |
| 2行 | 0 | 0 | 0 | $\frac{1}{2}$ | 1 | $-\frac{1}{2}$ | 4 |
| 3行 | 0 | 0 | 1 | $-\frac{1}{2}$ | 0 | $\frac{1}{2}$ | 4 |

我们准备进行另一个旋转步骤,但因表III中的0行不再含负的表值,进一步旋转无利可图。为了解这个事实,我们把0行转化成方程: 682

$$\pi = 760 - 25s_1 - 15s_3$$

在这个方程中,为使$\pi$最大,我们必须令$s_1 = s_3 = 0$;但解$S_3$恰是这样得到的。因此,$S_3$必为最优解!不出所料,这个最优解与在19.1节中用图解法所得到的结果是一致的。

根据图19.6,运用单纯形法可使我们有次序地从原点(初始极点)移到下一个极点(16.0),接着移到最优极点(16.4)。注意,我们已得出最优解,不必把五个极点都进行比较。还要注意,按照边际利润率标准,在表I中我们选择了$x_1$作为主列,经过较短的路程移到最优解;如果我们选$x_2$作为主元列,那么我们将首先移动图19.6中的点(0.8),需要三个旋转步骤才能到达点(16,4)。

**练习 19.5**

---

加上适当的虚拟变量,用单纯形法解下列线性规划:

1 Max. $\pi = 4x_1 + 3x_2$

s.t. $\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leqslant \begin{bmatrix} 4 \\ 6 \end{bmatrix}$

且 $x_1, x_2 \geqslant 0$

2 Max. $\pi = 4x_1 + 3x_2$

s.t. $\begin{bmatrix} 1 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leqslant \begin{bmatrix} 5 \\ 12 \end{bmatrix}$

且 $x_1, x_2 \geqslant 0$

3 Max. $\pi = 6x_1 + 2x_2 + 5x_3$

s.t. $\begin{bmatrix} 2 & 3 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \leqslant \begin{bmatrix} 10 \\ 8 \\ 19 \end{bmatrix}$

且: $x_1, x_2, x_3 \geqslant 0$

4 Max. $\pi = 3x_1 + 2x_2 + 6x_3$

满足上一问题的同样约束。

---

## 19.6 单纯形法的进一步说明

上面所讨论的极大化问题中,不需要寻找初始的 BFS,因为在原点总可得到一个现成的基可行解。但对极小化问题,因为原点683在可行区域的外面,上述事实就不成立了。

为便于说明,我们把 19.1 节的饮食问题转换成下列形式:

$$\text{Min.} \quad C = 0.6x_1 + x_2$$

$$\text{s.t.} \quad \begin{bmatrix} 10 & 4 & -1 & 0 & 0 \\ 5 & 5 & 0 & -1 & 0 \\ 2 & 6 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} 20 \\ 20 \\ 12 \end{bmatrix} \qquad (19.18)$$

$$\text{且} \quad x_1, x_2, s_1, s_2, s_3 \geqslant 0$$

注意，(非负)变量 $s_i$ 表示剩余(而非松弛)，必须从约束方程左边减去 $s_i$。从而，(19.18)中的系数矩阵后三列出现负的单位矩阵。如果我们试令 $x_1 = x_2 = 0$，三个约束条件产生解$(s_1, s_2, s_3,) = (-20, -20, -12)$，显然，它不是可行解。因此，我们必须积极寻找合适的基可行解。

## 在极小化问题中使用人工变量

简化这种搜寻的方法之一，是对每个约束条件加上非负人工变量把规划放大。人工变量用 $v_i$ 表示。在目标函数中，相对于选择变量的系数而言，人工变量的系数被赋予很大的值，原因后面再解释。若我们假设这些系数值为 100，那么(19.18)中的规划可写成下述形式：[①]

$$\text{Min.} \quad C = 0.6x_1 + x_2 + 100(v_1 + v_2 + v_3)$$

① 我们在 $3 \times 8$ 系数矩阵中加入了两条虚线，以明确三组列向量与三组变量(选择变量、剩余变量和人工变量)之间的联系。

$$
\text{s.t.}\quad \left[\begin{array}{cc:ccc:ccc} 10 & 4 & -1 & 0 & 0 & 1 & 0 & 0 \\ 5 & 5 & 0 & -1 & 0 & 0 & 1 & 0 \\ 2 & 6 & 0 & 0 & -1 & 0 & 0 & 1 \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ s_1 \\ s_2 \\ s_3 \\ v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 20 \\ 20 \\ 12 \end{bmatrix} \tag{19.18'}
$$

且 $x_1, x_2, s_1, s_2, s_3, v_1, v_2, v_3 \geqslant 0$

因为人工变量与三个线性无关的单位向量(组成单位矩阵)有联684系,与表(19.4)中的变量 $s_i$ 类似,现在我们不必去寻找初始基,可采用变量 $v$ 作为初始基。

尽管从一个初始基可行解出发是可以接受的,但我们必须用某种方法防止变量 $v_i$(引入变量 $v_i$ 仅是数学技巧)进入最优解中。如果在目标函数中给出变量 $v$ 系数很大,就不必对此感到担心,因为这些系数在本例(100)中表示某种假设的可接受的食品价格,就像 0.6 表示食品 I 的价格一样,这样对它们指定异常大的数值,自然,就使这些假设的食物太昂贵,而不能包含在最优食物中。于是,在最优解中必有 $v_1 = v_2 = v_3 = 0$,这样,就使增广线性规划(19.18′)最优等价于原规划(19.18)的最优。

现在可应用单纯形法了,其方式大体上与前面相同,只是稍作一些修改。首先,把(19.18′)中的规划整理成表(19.7)中的表 I,我们看到 $v_i$ 列还不是 4 - 元素单位向量的形式。为了把表 I 变成

表 II,首先我们把 100×(1 行 + 2 行 + 3 行)加到 0 行。因为表 II 的确含有 4×4 的单位矩阵,现在我们能读出解$(C, v_1, v_2, v_3) = (5200, 20, 20, 12)$。

另一个修改是关于选定主元列的标准。对于极小化的规划,我们应寻找 0 行中有最大正元素的列($C$ 列和常数列不予考虑)。当把表 II 的 0 行变成如下方程时,

$$C = 5200 - \frac{8497}{5}x_1 - 1499x_2 + 100(s_1 + s_2 + s_3)$$

采用上述新标准的原因也就容易看出了。如果在五个变量中选择一个作为进入变量以替换初始基中的一个人工变量,那么因为 $x_1$ 有绝对值最大的负系数,能使成本最小。但根据表 II,$x_1$ 是 0 行中具有最大正系数和变量,因此符合上述标准。因而应选 $x_1$ 列为主元列。

主元行的选定也严格按照前面的方法进行。这里因为最小的位移商是

$$\min\{\frac{20}{10}, \frac{20}{5}, \frac{12}{2}\} = \min\{2, 4, 6\} = 2$$

所以 1 行为主元行,且(围有虚线框)元素 10 是主元素。把主元列变成单位向量,我们最后得出表 III,表 III 给出解$(C, x_1, v_2, v_3) = (9006/5, 2, 10, 8)$,$x_1$ 已替换基向量 $v_1$,饮食成本大幅度下降:由 5200 美元降至 1800 美元稍多一点。

下一步只不过是重复前面的过程。但要注意,还需要两个旋转步骤才能抛弃剩下的两个人工变量。因为引进变量 $v_i$,不可避免地使计算过程变长了,所以我们应力求尽可能减少附加的人工变量个数。例如,如果系数矩阵的第 I 列恰巧含有元素 0,1 和 0

(而不是 10,5,2),那么因为这时可略去 $v_2$。在初始基中让 $x_1$ 代替它。同样,如果这列的元素是 0,0 和 1,我们可以省略 $v_3$。

685

**表 19.7 单纯形表 I 至表 VI**

| | | 选择变量 | | | 剩余变量 | | | 人工变量 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 表 | 行 | $C$ | $x_1$ | $x_2$ | $s_1$ | $s_2$ | $s_3$ | $v_1$ | $v_2$ | $v_3$ | 常数 |
| I | 0 | 1 | $-\frac{6}{10}$ | $-1$ | 0 | 0 | 0 | $-100$ | $-100$ | $-100$ | 0 |
| | 1 | 0 | 10 | 4 | $-1$ | 0 | 0 | 1 | 0 | 0 | 20 |
| | 2 | 0 | 5 | 5 | 0 | $-1$ | 0 | 0 | 1 | 0 | 20 |
| | 3 | 0 | 2 | 6 | 0 | 0 | $-1$ | 0 | 0 | 1 | 12 |
| II | 0 | 1 | $\frac{8497}{5}$ | 1499 | $-100$ | $-100$ | $-100$ | 0 | 0 | 0 | $\boxed{5200}$ |
| | 1 | 0 | 10 | 4 | $-1$ | 0 | 0 | 1 | 0 | 0 | 20 |
| | 2 | 0 | 5 | 5 | 0 | $-1$ | 0 | 0 | 1 | 0 | 20 |
| | 3 | 0 | 2 | 6 | 0 | 0 | $-1$ | 0 | 0 | 1 | 12 |
| III | 0 | 1 | 0 | $\frac{20481}{25}$ | $\frac{3497}{50}$ | $-100$ | $-100$ | $-\frac{8497}{50}$ | 0 | 0 | $\boxed{\frac{9006}{5}}$ |
| | 1 | 0 | 1 | $\frac{2}{5}$ | $-\frac{1}{10}$ | 0 | 0 | $\frac{1}{10}$ | 0 | 0 | 2 |
| | 2 | 0 | 0 | 3 | $\frac{1}{2}$ | $-1$ | 0 | $-\frac{1}{2}$ | 1 | 0 | 10 |
| | 3 | 0 | 0 | $\frac{26}{5}$ | $\frac{1}{5}$ | 0 | $-1$ | $-\frac{1}{5}$ | 0 | 1 | 8 |
| IV | 0 | 1 | 0 | 0 | $\frac{2498}{65}$ | $-100$ | $\frac{7481}{130}$ | $-\frac{8998}{65}$ | 0 | $-\frac{20481}{130}$ | $\boxed{\frac{35154}{65}}$ |
| | 1 | 0 | 1 | 0 | $-\frac{3}{26}$ | 0 | $\frac{1}{13}$ | $\frac{3}{26}$ | 0 | $-\frac{1}{13}$ | $\frac{18}{13}$ |
| | 2 | 0 | 0 | 0 | $\frac{5}{13}$ | $-1$ | $\frac{15}{26}$ | $-\frac{5}{13}$ | 1 | $-\frac{15}{26}$ | $\frac{70}{13}$ |
| | 3 | 0 | 0 | 1 | $\frac{1}{26}$ | 0 | $-\frac{5}{26}$ | $-\frac{1}{26}$ | 0 | $\frac{5}{26}$ | $\frac{20}{13}$ |
| V | 0 | 1 | 0 | 0 | $\frac{1}{15}$ | $-\frac{19}{75}$ | 0 | $-\frac{1501}{15}$ | $-\frac{7481}{75}$ | $-100$ | $\boxed{\frac{56}{15}}$ |
| | 1 | 0 | 1 | 0 | $-\frac{1}{6}$ | $\frac{2}{15}$ | 0 | $\frac{1}{6}$ | $-\frac{2}{15}$ | 0 | $\frac{2}{3}$ |
| | 2 | 0 | 0 | 0 | $\frac{2}{3}$ | $-\frac{26}{15}$ | 1 | $-\frac{2}{3}$ | $\frac{26}{15}$ | $-1$ | $\frac{28}{3}$ |
| | 3 | 0 | 0 | 1 | $\frac{1}{6}$ | $-\frac{1}{3}$ | 0 | $-\frac{1}{6}$ | $\frac{1}{3}$ | 0 | $\frac{10}{3}$ |

| | | 选择变量 | | | 剩余变量 | | | 人工变量 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 表 | 行 | $C$ | $x_1$ | $x_2$ | $s_1$ | $s_2$ | $s_3$ | $v_1$ | $v_2$ | $v_3$ | 常数 |
| VI | 0 | 1 | 0 | 0 | 0 | $-\frac{2}{25}$ | $-\frac{1}{10}$ | $-100$ | $-\frac{2498}{25}$ | $-\frac{999}{10}$ | $\boxed{\frac{14}{5}}$ |
| | 1 | 0 | 1 | 0 | 0 | $-\frac{3}{10}$ | $\frac{1}{4}$ | 0 | $\frac{3}{10}$ | $-\frac{1}{4}$ | 3 |
| | 2 | 0 | 0 | 0 | 1 | $-\frac{13}{5}$ | $\frac{3}{2}$ | $-1$ | $\frac{13}{5}$ | $-\frac{3}{2}$ | 14 |
| | 3 | 0 | 0 | 1 | 0 | $\frac{1}{10}$ | $-\frac{1}{4}$ | 0 | $-\frac{1}{10}$ | $\frac{1}{4}$ | 1 |

在表(19.7)中从表 II 开始,后面的每一表都说明成本减少。在表 686 VI 中,成本减至 14/5 = 2.80 时,从 0 行可看出它不可能进一步减少了,因为在 $x_i$,$s_i$ 和 $v_i$ 列中再也没有正的表值了。因此最优解为

$$(\bar{x}_1, \bar{x}_2, \bar{s}_1, \bar{s}_2, \bar{s}_3) = (3, 1, 14, 0, 0)$$

$$\bar{C} = \$ 2.80$$

这同在 19.1 节用图解法得到的结果是一致的。

## 人工变量的另一个应用

人工变量并不专门用于极小化问题。在某些极大化的规划问题中,也具有其优势。

在求极大值的情况下,若其中一个(比如,第三个)约束条件恰巧是严格等式,那就不需要 $s_3$ 变量了。如果是这种情况,我们就从单纯形表中"取走"一个单位向量,所谓"现成的"初始基可行解再也不能得到了。为了解决这个问题,我们让人工变量 $v_3$(对第三个约束条件)填补由于缺乏 $s$ 变量而留下的空位,因为它也能产生适当类型的单位向量。

当然,为了确保 $v_i$ 能正常的排斥在最优解之外,在这种情况下,我们必须赋予它在目标函数中的系数以数字很大的负值(负边

际利率)。但在其他方面,应用单纯形法与前面完全一样。

## 退　化

前面所讨论的线性规划有一共性:$m$ 个约束方程中的常数向量不能表示为小于 $m$ 个系数向量的线性组合。例如,在(19.11′)中右边向量不能表示为小于三个左边向量的线性组合,因此,其中 $m$ 个变量的每个都必取非零值;这就是为什么在表 19.3 的解空间中,每点恰有三个非零元素的原因。当这个性质不成立时,我们就说这个线性规划是退化的。

就单纯形法而言,退化唯一的表现形式在于"相同的"位移商(tied displacement quotient)出现。即两个或两个以上的商都具有最小的特征,因而两行或更多的行同样都是主元行的候选行。由于一次不可能同时替换一个以上的变量,必须寻找某个判据,以摆脱最小位移商相同的情况。

687　如果是任意选择的话,一个实用的方法,就是这样选取主元行:在相同位移商的变量中,实际上开始被替换的是位于单纯形表中最左边的变量。有了这个摆脱最小位移商相同的判据,那末旋转步骤剩下的工作就可有次序地进行。

在退化情况下,旋转步骤可能不会提高利润或减少成本。在迭代过程得以打破僵局以前,采取几个旋转步骤实际上可能并没有改善。单纯形法应用于退化情况,除了相同的位移商和没有改善的可能性以外,在其他方面同以前是一样的。

单纯形法或单纯算法[①] 就介绍到这里。这种方法本身并不

---

① "算法"是一个新词,意指程序化的计算方法。

难,但在高维线性规划中,计算工作冗长而又乏味。幸运的是,现代计算机能很好的适应线性规划所需要的重复计算。适当地给计算机一组详细指令(即**计算机的程序设计**)时,我们可依靠机器完成单纯形算法。这种计算可靠,持久,且速度远远超过人工计算。这样,高维线性规划的计算问题仅仅受计算机的计算能力所限,几乎不成什么问题了。

## 练习 19.6

1 完成自表 19.7 中从表 II 到后面每一表的计算。

用单纯形算解下列各题:

2 Min. $C=x_1+4x_2$

s.t. $\begin{bmatrix}1 & 2\\3 & 2\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}\geqslant\begin{bmatrix}8\\12\end{bmatrix}$

且 $x_1,\ x_2\geqslant 0$

3 Min. $C=2x_1+7x_2$

s.t. $\begin{bmatrix}1 & 2\\0 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}\geqslant\begin{bmatrix}8\\3\end{bmatrix}$

且 $x_1,\ x_2\geqslant 0$

[提示:仅需一个人工变量]

4 Min. $C=12x_1+42x_2$

s.t. $\begin{bmatrix}1 & 2\\1 & 4\\3 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}\geqslant\begin{bmatrix}3\\4\\3\end{bmatrix}$

且 $x_1,\ x_2\geqslant 0$

# 第 20 章　线性规划(续)

## 20.1　对偶性

到目前为止,我们一直将极大化和极小化的线性规划问题作为两类独立的问题来讨论。但实际上对应于每个极小化的规划(使 $C$ 极小),总存在对应的极大化规划(使新变量 $C^*$ 极大),而且具有这一性质,即 $\bar{C}^* = \bar{C}$。类似地,每一个$\pi$极大化规划,总是对应着一个 $\pi^*$ 极小化规划,使得$\bar{\pi}^*=\bar{\pi}$。原来的线性规划一般称为**原规划**,其对应规划称为**对偶规划**(或简称**对偶**)。由此看来,极大化和极小化规划,实际上并不是像它们的含义表示的那样没有联系。事实上,在原规划和对偶规划中,目标函数的最优值总是相等的。这样,我们就有在两个规划中挑选较容易的一个进行研究的余地;正如我们将看到的,总可以把对偶规划的变量解值转化为原规划变量解值,反之亦然。

### 对偶规划

为明确区分起见,我们用 $x_j$ 表示原规划的选择变量(像前面那样),用 $y_i$ 表示对偶规划的选择变量。那么,原规划和对偶规划结构的相互关系,可用下面两个例子加以说明。

**例 1** 原规划 对偶规划

原规划：

$$\text{Max.}\quad \pi = 3x_1 + 4x_2 + 3x_3$$

$$\text{s.t.}\quad \begin{bmatrix}1 & 1 & 3\\ 2 & 4 & 1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} \leqslant \begin{bmatrix}12\\ 42\end{bmatrix}$$

且 $x_1, x_2, x_3 \geqslant 0$

对偶规划：

$$\text{Min.}\quad \pi^* = 12y_1 + 42y_2$$

$$\text{s.t.}\quad \begin{bmatrix}1 & 2\\ 1 & 4\\ 3 & 1\end{bmatrix}\begin{bmatrix}y_1\\ y_2\end{bmatrix} \geqslant \begin{bmatrix}3\\ 4\\ 3\end{bmatrix}$$

且 $y_1, y_2 \geqslant 0$

**例 2**

原规划：

$$\text{Min.}\quad C = 4x_1 + 3x_2 + 8x_3$$

$$\text{s.t.}\quad \begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} \geqslant \begin{bmatrix}2\\ 5\end{bmatrix}$$

且 $x_1, x_2, x_3 \geqslant 0$

对偶规划：

$$\text{Max.}\quad C^* = 2y_1 + 5y_2$$

$$\text{s.t.}\quad \begin{bmatrix}1 & 0\\ 0 & 1\\ 1 & 2\end{bmatrix}\begin{bmatrix}y_1\\ y_2\end{bmatrix} \leqslant \begin{bmatrix}4\\ 3\\ 8\end{bmatrix}$$

且 $y_1, y_2 \geqslant 0$

简单地说，变换法则是：(1)把“求极大”改为“求极小”，反之亦然。(给原规划求极大或求极小的目标函数标上星号，使之成为对偶规划求极小或求极大的目标函数。)(2) 原约束条件中的不等式符号，在对偶约束条件中必反号，虽然非负限制中的不变。(3)取原约束条件的系数矩阵的转置，作为对偶约束条件的系数矩阵。原目标函数中的系数行向量转置后，视为对偶约束条件中的常数列向量。同样，原条件中的常数列向量转置后，成为对偶目标函数中的系数行向量。

根据这些变换规则读者应很容易推出：如果我们把 $y$ 转换为 $x$，并在目标函数中去掉(而不是附上)星号，那么，对偶规划的对偶就是原规划本身。

为了从上面所举的两个例子中引出一般性结论，我们用矩阵

符号把原规划和对偶规划关系表示如下：

| 原规划 | | | 对偶规划 | | |
|---|---|---|---|---|---|
| Max. | $\pi=c'x$ | | Min. | $\pi^{*}=r'y$ | |
| s.t. | $Ax\leqslant r$ | $\rightarrow$ | s.t. | $A'y\geqslant c$ | (20.1) |
| 且 | $x\geqslant 0$ | | 且 | $y\geqslant 0$ | |

和

| | | | | | |
|---|---|---|---|---|---|
| Min. | $C=c'x$ | | Max. | $C^{*}=r'y$ | |
| s.t. | $Ax\geqslant r$ | $\rightarrow$ | s.t. | $A'y\leqslant c$ | (20.2) |
| 且 | $x\geqslant 0$ | | 且 | $y\geqslant 0$ | |

如果原规划有 $m$ 个约束条件，$n$ 个选择变量，因而矩阵 $A$ 为 $m\times n$ 维矩阵，那么对偶规划将有 $n$ 个约束条件，$m$ 个选择变量，因为矩阵 $A'$是 $A$ 的转置，$A'$为 $n\times m$ 维矩阵。原规划中其它矩阵的维数如下：$c'$是 $1\times n$ 维，$x$ 是 $n\times 1$ 维，$r$ 是 $m\times 1$ 维，正如(19.5)所示。对偶规划中的向量 $y$ 为 $m\times 1$ 维。

690

原规划和对偶规划的解法难易程度不同，其原因在于它们所含的约束条件和选择变量数可能不同。在上面的两个例子中，原规划(含 3 个变量)用图解法不易解出，其对偶规划只含两个变量，这就很适合图解法。而且，即使原规划和对偶规划都不能用图解法解，因而无论如何必须求助于单纯形算法时，我们仍然希望选择那些约束条件较少的规划求解。因为约束条件个数越少，基的维数就越少，我们必须增添的虚拟变量也就越少。最后，即使原规划和对偶规划的约束条件个数相等(或接近相等)，我们还是会发现选取极大化规划比较容易，因为后者通常含有现成的初始基可行解，因而可不必为人工变量费心。

## 对偶定理

虽然线性规划的上述解法具有一定程度的灵活性，但这种自由选择也产生一个问题：如果我们决定解对偶规划，其解包括 $\bar{y}_i$（比方说与 $\bar{C}^*$）的值，但如我们真正关心的是规划的解值 $x$（与 $\bar{C}$），那么，我们怎样由对偶规划的解求出原规划的解，或者由原规划求出对偶规划的解呢？这些问题的答案在于下面两个定理：

对偶定理 I：如果最优可行解存在，则原规划和对偶规划目标函数的最优值总是相等的，即：$\bar{C} = \bar{C}^*$ 和$\pi = \bar{\pi}^*$。

对偶定理 II：($a$)如果线性规划中某一选择变量的最优值为非零，那么对偶规划中对应的虚拟变量的最优值必为零值。即（如用 $s_i$ 表示原规划的第 $i$ 个虚拟变量，$t_j$ 表示对偶规划的第 $j$ 个虚拟变量）：

$$\bar{y}_i > 0 \Rightarrow \bar{s}_i = 0 \quad 和 \quad \bar{x}_j > 0 \Rightarrow \bar{t}_j = 0$$

($b$)如果线性规划中某个虚拟变量的最优值非零，那么对偶规划对应的选择变量的最优值必等于零。即

$$\bar{s}_i > 0 \Rightarrow \bar{y}_i = 0 \quad 和 \quad \bar{t}_j > 0 \Rightarrow \bar{x}_j = 0$$

按照第一个定理，就目标函数的最优值而论，原规划与对偶规划之间的选择是无关紧要的。第二个定理阐明了通常所说的一个规划的选择变量与另一个规划的虚拟变量之间的互补松弛关系。借助第二个定理，已知一线性规划的最优解将为计算其对偶规划的最优解提供了充分的信息。 691

在介绍这两个定理的应用之前，概述一下这两个定理的合理性也许是重要的。我们将具体考察最大化形式的原规划，但推理过程对最小化形式也是类似的。最大化原规划的约束可以像

(20.1)一样，用矩阵不等式 $Ax \leqslant x$ 表示。使用松弛变量 $s$ 的向量，我们可以将不等式转化为矩阵方程：

$$Ax + s = r \qquad (s \geqslant 0) \tag{20.3}$$

如果问题包含 $m$ 个原规划约束，$n$ 个原规划选择变量，则 $A$ 为 $m \times n$ 维矩阵，$x$ 是 $n \times 1$ 维的，从而 $Ax$（像 $s$ 和 $r$ 一样）是 $m \times 1$ 维的。因此，以 $1 \times m$ 向量左乘(20.3)中的每一项是可行的。因对偶规划应具有 $m$ 个选择变量，向量 $y$ 的维数是 $m \times 1$，其转置 $y'$ 是 $1 \times m$。当用 $1 \times m$ 向量左乘(20.3)的每一项时，得到

$$\begin{aligned} y'Ax + y's &= y'r \\ &= r'y = \pi^* \qquad [\text{由(20.1)}] \end{aligned} \tag{20.4}$$

注意，(20.4)中的每一项是 $1 \times 1$ 的，即是一个标量。两个表达式 $y'r$ 和 $r'y$ 不仅本质上是标量积，而且在数值上也是相等的。

现在我们转向对偶规划，我们可以像在(20.1)中那样，用矩阵不等式 $A'y \geqslant c$ 来表示其约束。但是通过使用剩余变量（我们将以 $t$ 表示之），可以将其转化为矩阵方程：

$$A'y - t = c \qquad (t \geqslant 0) \tag{20.5}$$

因(20.5)中的每一项的维数为 $n \times 1$，所以，我们可以用 $1 \times n$ 向量 $x'$ 左乘它的每一项，得到结果

$$x'A'y - x't = x'c$$

或者，取每一项的转置，

$$\begin{aligned} y'Ax - t'x &= c'x \qquad [\text{由(4.11)和练习 4.6 - 3}] \\ &= \pi \qquad [\text{由(20.1)}] \end{aligned} \tag{20.6}$$

由(20.4)和(20.6)，立即可以得到

$$\pi^* - \pi = y's + t'x \geqslant 0 \qquad [\text{由非负限制}] \tag{20.7}$$

因此，$\pi^*$（对偶规划被求极小值）永远也不会低于 $\pi$（原规划被求极

大值)。而且,$\pi$ 也永远不会超过$\pi^*$。但是,当$\pi^*$被极小化,$\pi$被极大化时,这两个值将是相等的。这构成了定理 I 结论$\overline{\pi}=\overline{\pi}^*$的基础。对定理 I,我们在 21.4 节还要进一步讨论。

在对偶规划和原规划的最优解中,$\overline{\pi}$ 与$\overline{\pi}^*$相等,(20.7)意味着 692

$$\bar{y}'\bar{s}+\bar{t}'\bar{x}=0$$

或更明确地,

$$(\bar{y}_1\bar{s}_1+\cdots+\bar{y}_m\bar{s}_m)+(\bar{t}_1\bar{t}_1+\cdots+\bar{t}_n\bar{x}_n)=0 \quad (20.8)$$

考虑到对变量的非负性限制,满足方程的唯一方式是使 $m+n$ 个乘积项的每一个均等于零。这又直接引出定理 II 的结论。注意到这一点是重要的:要求(20.8)中每个乘积项为零,使我们只能使用已知的乘积项中的一个因子为正(比如,$\bar{y}_i>0$),作为推断乘积中的另一个因子为零($\bar{s}_i=0$)的基础。它不允许我们取其中的一个因子为零(比如 $\bar{t}_j=0$),作为乘积中的另一个因子为正($\bar{x}_j>0$)这一结论正确性的保证。因为 $\bar{t}_j$ 为零与 $\bar{x}_j$ 为零可同时存在,并仍能满足 $\bar{t}_j\bar{x}_j=0$ 的要求。

## 通过对偶规划解原规划

现在让我们利用上面的例 1 来说明对偶定理的应用。因为该例的对偶规划只有两个选择变量,可用图解法解之。所以,我们先用图解法求出对偶规划的最优解,然后利用对偶定理求原规划的最优解。

读者可以验证:对偶规划的图解分析可得到如下信息:

$$(\bar{y}_1,\bar{y}_2)=(2,\frac{1}{2})$$

因而,由目标函数,容易算出

$$\bar{\pi}^* = 12(2) + 42(\frac{1}{2}) = 45$$

而且,把 $\bar{y}_i$ 值代入对偶规划的约束条件时,可看出头两个约束条件会简化为两个方程,而第三个约束条件简化为严格的不等式,于是 $\bar{t}_1 = \bar{t}_2 = 0$,而 $\bar{t}_3 > 0$。根据对偶规划的这些信息,我们能够求出原规划的最优解。

因 $\bar{t}_1$ 和 $\bar{y}_2$ 都不等于零,所以由对偶定理 II 中给出的互补松弛关系我们必有 $\bar{s}_1 = \bar{s}_2 = 0$,因而原规划的头两个(只有两个)约束条件成为严格等式,且因 $\bar{t}_3 > 0$,由此得 $\bar{x}_3 = 0$。这样在最优解的情况下,例 1 中原规划的约束条件部分表现为下述形式:

$$\begin{bmatrix} 1 & 1 & 3 \\ 2 & 4 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} = \begin{bmatrix} 12 \\ 42 \end{bmatrix} \quad \text{或} \quad \begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 12 \\ 42 \end{bmatrix} \tag{20.9}$$

693 由此我们得到 $\bar{x}_1 = 3$ 和 $\bar{x}_2 = 9$,因此,原规划中的选择变量的解值用等式表示为:

$$(\bar{x}_1, \bar{x}_2, \bar{x}_3) = (3, 9, 0) \tag{20.10}$$

这就完成了从 $\bar{y}_i$ 到 $\bar{x}_j$ 的转换。

根据对偶定理 I,$\bar{\pi}$ 值应为 45。如果我们将 $\bar{x}_j$ 值代入原规划的目标函数,实际上也就验算了答案:$\bar{\pi} = 3(3) + 4(9) + 0 = 45$。

关于线性规划的对偶性,特别值得注意的是下述重要的有价值的事实:如果用单纯计算法(而不是用图解法)解对偶规划,那么,从其最优(最后)表中,我们不仅可读出对偶规划的最优解,而且可直接读出原规划的最优解!这样,由 $\bar{y}_i$ 到 $\bar{x}_j$ 的转换现在实际上就是自然而然的事了。

**表 20.1　对偶规划最优表**

| 行 | $\pi^*$ | $y_1$ | $y_2$ | $t_1$ | $t_2$ | $t_3$ | $v_1$ | $v_2$ | $v_3$ | 常数 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | $-3$ | $-9$ | 0 | 不相关， | | | 45 |
| 1 | 0 | 0 | 0 | $-\frac{11}{2}$ | $\frac{5}{2}$ | 1 | 因而 | | | $\frac{7}{2}$ |
| 2 | 0 | 0 | 1 | $\frac{1}{2}$ | $-\frac{1}{2}$ | 0 | 可省略。 | | | $\frac{1}{2}$ |
| 3 | 0 | 1 | 0 | $-2$ | 1 | 0 | | | | 2 |

我们利用例 1 的对偶规划最优表(如表 20.1 所示)来加以说明。由表中的四个单位向量容易读出对偶的最优解为：

$$(\bar{\pi}^*, \bar{y}_1, \bar{y}_2, \bar{t}_3) = (45, 2, \frac{1}{2}, \frac{7}{2})$$

此处 $t_3$ 表示第三个对偶虚拟变量。

当然,这四个数字都取自常数列。但如果我们看到 0 行中虚变量列的元素取矩形中的三个表值的绝对值,我们得到 3,9 和 0 三个数,排成适当的次序恰得(20.10)中的 $\bar{x}_1$, $\bar{x}_2$ 和 $\bar{x}_3$ 的值。[①] 通常,我们总是把原规划最优表中的 $t_j$ 列与变量 $x_j$ 联系起来,并将 $t_j$ 列中第一个元素的绝对值看作原规划选择变量的最优值。[②] 而且我们也能把 $y_i$ 列同原规划的虚拟变量联系起来,且将这些列中第一个元素的绝对值看作 $\bar{s}_i$ 的值。在表 20.1 中,0 行曲线框中的两个零元素表示 $\bar{s}_1 = \bar{s}_2 = 0$,我们早已知道这结果是正确的。694 此外,由对偶定理 I 知道常数列中标有方形的元素也决定值 $\bar{\pi}$。关于原规划最优解的有关数据都可由对偶规划最优表的 0 行读数

① 取绝对值是必要的,因为 $\bar{x}_j$ 必为非负值。

② 因对偶规划中的虚拟变量个数同原规划中的选择变量的个数相等,所以在这个过程中,不会产生维数问题。

推断出来！同样，已知原规划的最优表也可由其0行读出对偶规划的最优解。

这些极易得到的结果实质上与对偶定理II完全相符。最优解中取非零值的任意对偶变量(此处为 $y_1$, $y_2$，和 $t_3$)，其特点是在最优表中必是单位列向量。因为$\pi^*$列已先取 $e_1$ 类型的单位向量(单位元素在最上面)，因而在 $y_1$, $y_2$ 和 $t_3$ 列中其单位向量的第1个元素必为零值。按照刚才详细说明的方法，我们可推出$\bar{s}_1 = \bar{s}_2 = \bar{x}_3 = 0$。而这个结论恰是对偶定理II要说明的。同样，按照相反的次序应用这种方法，$\bar{x}_1$ 和 $\bar{x}_2$ 非零(在原规划的最优表中)就意味着 $\bar{t}_1$ 和 $\bar{t}_2$ 必为零；这就防止了表20.1中的 $t_1$ 和 $t_2$ 列化为单位向量。

## 练习 20.1

---

已知下列原规划，写出其对偶规划的表达式：

1 Max. $\pi = 9x_1 + x_2$

s.t. $2x_1 + x_2 \leqslant 8$

$4x_1 + 3x_2 \leqslant 14$

且 $x_1, x_2 \geqslant 0$

2 Min. $C = x_1 + 7x_2$

s.t. $\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \geqslant \begin{bmatrix} 5 \\ 4 \\ 9 \end{bmatrix}$

且 $x_1, x_2 \geqslant 0$

3 Max. $\pi = c_1x_1 + c_2x_2 + c_3x_3$

$$\text{s.t.} \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \leqslant \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}$$

且 $\quad x_j \geqslant 0 \ (j = 1, 2, 3)$

4　课文中例 2 的对偶规划，其最优表如下：

| 行 | $C^*$ | $y_1$ | $y_2$ | $t_1$ | $t_2$ | $t_3$ | 常数 |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 1 | 2 | 19 |
| 1 | 0 | 0 | 0 | 1 | 2 | −1 | 2 |
| 2 | 0 | 0 | 1 | 0 | 1 | 0 | 3 |
| 3 | 0 | 1 | 0 | 0 | −2 | 1 | 2 |

($a$) 写出原规划和对偶规划的最优解。

695

($b$) 将 $y$ 和 $x$ 分别代入对偶规划和原规划的目标函数，验证 $C^* = C$。

5　结过若干个旋转步骤后，已知的线性规划表为下列形式：

| 行 | $\pi$ | $x_1$ | $x_2$ | $s_1$ | $s_2$ | 常数 |
|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 2 | 1 | 14 |
| 1 | 0 | 0 | 1 | 2 | −1 | 2 |
| 2 | 0 | 1 | 0 | −1 | 1 | 2 |

($a$) 这是一个最优表吗？为什么？

($b$) $\bar{x}_1$，$\bar{x}_2$ 和$\bar{\pi}$ 的值等于多少？

($c$) 其对偶规划应含多少个选择变量？为什么？

($d$) 对偶选择变量及其目标函数和最优值等于多少？

6　已知下列线性规划：

Min. $\quad C = x_1 + 4x_2$

$$\text{s.t.} \quad \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \geqslant \begin{bmatrix} 8 \\ 12 \end{bmatrix}$$

且 $\quad x_1, \ x_2 \geqslant 0$

的最优解是 $\bar{x}_1 = 8$ 和 $\bar{x}_2 = 0$：

($a$) 求 $\bar{s}_1, \bar{s}_2$ 和 $\bar{C}$；

($b$) 写出对偶规划；

($c$) 根据上面($a$)小题的答案，$\bar{y}_2$ 的值应等于多少？根据已给的最优解，哪一个对偶约束条件是严格等式？

($d$) 用其($c$)小题的答案解出 $\bar{y}_1$。

($e$) 将 $\bar{y}_1$ 和 $\bar{y}_2$ 代入函数 $C^*$,验算是否有 $\bar{C}^* = \bar{C}$。

7　在(20.1)和(20.2)中的从原规划到对偶规划的变换法则仅能处理约束条件全为不等式的情况。现在,考察一个具有等式约束的线性规划:

$$\begin{aligned} \text{Min.} \quad & C = c_1x_1 + c_2x_2 + c_3x_3 \\ \text{s.t.} \quad & a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \geqslant r_1 \\ & a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = r_2 \\ \text{且} \quad & x_1,\ x_2,\ x_3 \geqslant 0 \end{aligned}$$

($a$) 设计一种写出此线性规划的方法,使得我们可以像平常一样应用(20.2)。写出对偶规划。对偶规划与原规划的选择变量数一样吗?

($b$) 证明:通过引入一个新的变量 $y_4 = y_2 - y_3$,我们可以恢复对偶规划选择变量数和原规划的选择变量数之间的一致性。$y_4$ 为负吗?为什么是?或者为什么不是?

($c$) 评价下列陈述的正确性:“即便原规划中的第 $j$ 个约束恰好为严格等式,我们仍然可以应用构建对偶规划的一般的法则,但第 $j$ 个对偶选择变量的符号不应受到限制这一点除外”。

---

696

## 20.2　对偶的经济解释

现在我们已清楚,线性规划的对偶规划可用来代替原规划的计算。但在绝大多数情况下,对偶规划本身也具有重要的经济意义。

### 生产规划的对偶规划

我们首先解释生产问题的对偶规划。在两种产品、两个约束条件的简单情况下,原规划取下列形式:

$$\text{Max.} \quad \pi = c_1x_1 + c_2x_2$$

$$\text{s.t.} \quad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leqslant \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}$$

且　　$x_1, x_2 \geqslant 0$

因而,对偶规划可写成:

$$\text{Max.} \quad \pi^* = r_1 y_1 + r_2 y_2$$

$$\text{s.t.} \quad \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \geqslant \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

且　　$y_1, y_2 \geqslant 0$

在解释对偶规划以前,我们必须首先明确对偶变量(包括其量度单位)的一般性质。在原规划中$\pi$表示以美元为单位的总利润额。[①] 由于$\pi^* = \pi$,对偶规划中的解$\pi^*$也应以美元为单位,就像表达式 $r_1 y_1 + r_2 y_2$ 一样。因为符号 $r_i$ 是指厂商固定工厂中第 $i$ 种资源的总量(在以前的例子中称为第 $i$ 个生产间的能力),显然符号 $y_i$ 表示 $i$ 种资源每单位的美元数,只有这时每项 $r_i y_i$ 才是美元数。这就是说,$y_i$ 表示讨论中的某种资源估算的价值。

显然,这种价值不是市场价格;这种资源早已为厂商所拥有,并成为厂商固定成本的一部分,而且,厂商也并没有在市场中以 $y$ 价格来购买它。更确切地说,它是这种资源的估算价值。因此,$y_i$ 值称为第 $i$ 种资源的会计价格(accounting price)或影子价格。我们认为把 $y_i$ 看作使用第 $j$ 种资源时的机会成本是合适的。

由此,我们来考察对偶规划的三种成分。第一,非负限制 $y_i \geqslant 0$ 表示任意一种资源的估算价值不允许小于零。当然,这是经济上的合理要求。事实上,我们总把资源的价值估算为正值,除非这种资源恰巧不被充分利用,因而资源在生产使用中产生零机会成 697

① 像在 19.1 节解释的那样,总利润在这里表示尚未扣除任何固定成本,但扣去可变成本的总销售收益。

本。这意味着一种资源的正机会成本($\bar{y}_i>0$)总与最优解($s_i=0$)中充分利用这种资源相联系。读者注意,根据这种解释,我们再回到第二个对偶定理(在此以前仅出它的数学解释),现在它获得了经济含义。

其次,再转到对偶*约束条件*,我们可具体地考察(20.11)中的第一个约束条件:

$$a_{11}y_1+a_{21}y_2\geqslant c_1 \tag{20.12}$$

因为系数 $a_{ij}$ 表示生产第 $j$ 种单位产品所使用的第 $i$ 种资源的数量,所以(20.12)的左边表示生产一个单位第 1 种($j=1$)产品的总机会成本。右边的 $c_1$ 项表示每单位第一种产品的总利润。于是这个约束条件是要求所估算的生产机会成本至少与产品的总利润一样大。

现在,我们稍微思考一下,如果生产的机会成本实际上超过利润,那么资源的配置一定不是最优的,因为简单地放弃第一种产品的生产时,从那里节约的资源可立即在其它地方更好地获得利用。在数学上,如果(20.12)中的≥符号中>部分在最优解中成立,那么第一种产品就不应生产($x_1=0$)。反之,如果第一种产品实际上在生产时,即如果 $\bar{x}_1\neq0$ 时,那么生产的机会成本必恰好等于总利润,且(20.12)中≥符号中的=部分在最优解中成立。简言之,$\bar{t}_1>0\Rightarrow x_1=0$,且 $\bar{x}_1>0\Rightarrow\bar{t}_1=0$。但这仅仅是复述对偶定理 II 而已。当然(20.11)中的第二个约束条件也可给同类似解释。

最后,让我们看看对偶*目标函数*。当厂商的固定工厂中可利用的两种资源总数为 $r_1$ 和 $r_2$ 时,显然表达式

$$\pi^*=r_1y_1+r_2y_2$$

表示这些资源的总估算价值。满足上面所解释的约束条件,使这

个总值极小就是对偶规划的目标。于是，如何生产每种产品的机会成本不小于该产品的总利润，那么原规划和对偶规划之间的对应就表示：求出最优产出水平使利润极大（原规划）就等价于该厂资源的总结算值或机会成本极小。$\bar{\pi}=\bar{\pi}^*$意指在最优解的情况下，通过影子价格将该厂的各种资源作为一个整体来估算或确定其总利润。对偶规划的经济意义就解释到这里。

为简单起见，我们就两种产品两种资源的简单模型考察了生 698 产问题。但因前面讨论使用的推理与假设 $m=n=2$ 无关，所以对偶规划的机会成本的解释容易推广到 $m$ 种资源，$n$ 种产品的结构中去。

## 饮食问题的对偶规划

同理，饮食问题的对偶规划中的选择变量也可给出估算价值的解释。

用矩阵符号，一般饮食问题可表示为：

Max. $C=c'x$

s.t. $Ax\geqslant r$

且 $x\geqslant 0$

这里 $C$ 是饮食的总成本，$c'$是食物价格行向量，$x$ 是食品数量的列向量，$A$ 是系数 $a_{ij}$（第 $j$ 种食物中所含的第 $i$ 种营养数量）所组成的矩阵，$r$ 是最小营养需要列向量。因此，对偶规划问题可写成：

Max. $C^*=r'y$

s.t. $A'y\leqslant c$

且 $y\geqslant 0$

其中 $y$ 是对偶选择变量的列向量。

既然 $C$ 是以美元为单位，其对偶规划中对应的 $C^*$ 也必以美元为单位。但向量 $r'$ 的元素单位是各个营养成分的物质单位，因而，向量 $r$ 的元素也应以各个单位营养成分的美元数为单位。换言之，对偶选择变量又具有某种估算价值的性质。这一次，它们表示各种营养成分的估算价值。因此，不允许它们取负值。

就此而论，每个对偶约束条件可解释为每种单位食品所含营养成分的总估算价值应不大于该食品的价格。实际上，如果第 $j$ 种食品的价格超过其营养成分的估算价值，我们决不会购买这种食品，这是普通常识，因为我们花掉这笔钱是不合算的。因此，在 699 最优解的情况下，我们只能购买这样的食品，其各种营养成分的估算价值恰等于各种营养成分的价格。精明的读者将注意到这又是对偶定理 II 在起作用。

作为这个对偶规划的目的，我们设法使起码的营养需要量的总估算值极大化，且满足上面列出的约束条件。达到这个极大值，实际上也就使饮食成本最低，而又能满足起码的营养需要量。

## 对偶选择变量与拉格朗日乘数

已经证明对偶选择变量 $y$ 表示影子价格或估算价值。我们也能证明在最优解的情况下，它们在线性规划中所起的作用与古典最优化问题中拉格朗日乘数所起的作用一样。即它们都可用来度量原规划的目标函数最优值对原规划的约束常数变化的灵敏度。

为了证明这一点，我们还需考察生产问题以及(20.11)中说明

的其对偶问题。在最优解的情况下,变量 $y_i$ 取 $\bar{y}_i$,对偶规划中的目标函数成为$\bar{\pi}^* = r_1\bar{y}_1 + r_2\bar{y}_2$。但因,$\pi = \bar{\pi}^*$,也可写成

$$\bar{\pi} = r_1\bar{y}_1 + r_2\bar{y}_2$$

因此,对 $r_1$ 和 $r_2$ 微分$\bar{\pi}$,得到

$$\bar{y}_1 = \frac{\partial\bar{\pi}}{\partial r_1} \quad 和 \quad \bar{y}_2 = \frac{\partial\bar{\pi}}{\partial r_2}$$

这表示 $\bar{y}_i$(第 $i$ 个对偶选择变量的最优值)是$\bar{\pi}$(原规划的目标函数的最优值)对 $r_i$(原规划的第 $i$ 个约束条件中的参数常数)无限小变化的灵敏度的量度。具体地说,$\bar{y}_i$ 告诉我们第 $i$ 个生产能力约束条件稍为松弛一点(第 $i$ 种资源数量增加)时,在最优解的情况下如何改变总利润。因为在使 $z = f(x, y)$ 极大,且满足 $g(x, y) = c$ 的古典问题中,得到拉格朗日函数 $Z = f(x, y) + \lambda[c - g(x, y)]$,拉格朗日乘数的最优值具有下述含义:

$$\bar{\lambda} = \frac{d\bar{Z}}{dc} \qquad [见(12.16)]$$

很明显,对偶选择变量和拉格朗日乘数虽然结构不相同,但所起的作用是一样的。

当然,$\bar{y}_i$ 的拉格朗日乘数的解释与前面把 $y_i$ 解释为第 $i$ 种资源的估算价值是完全一致的。当第 $i$ 个约束条件松弛时,第 $i$ 种资源现有的数量增大,它对总利润的影响(拉格朗日乘数的解释)将直接取决于第 $i$ 种资源估算价值的大小(前面的解释),因为在最优解的情况下,总利润和厂商资源的总估算价值必彼此相等。 700

## 练习 20.2

---

1 练习 19.1－6 曾给出约会的规划问题，试列出其对偶规划的数学表达式，并给出其经济解释。

2 在饮食问题中，试给出对偶选择变量的最优值 $\bar{y}_i$ 的拉格朗日乘数解释。并具体地说明它们用来度量什么。这个解释与 $y_i$ 估算价值的解释一致吗？

---

# 20.3 活动分析：微观水平

到目前为止，在我们的讨论中，线性规划的约束条件部分一般都是按横行来读的，即每个约束不等式都看作一个整体，得到一边界线（或超平面），这边界线又产生两个半空间。与此同时，在讨论基本可行解方面我们也可改变上述作法，沿纵列读出的约束条件的系数，把它看成(19.11′)和(19.12)中的列向量，这种纵向观点提供了考察线性规划的另一种方法。

### 活动的概念

设想一厂商为生产两种物品$(x_1, x_2)$，使用两种资源（$K$ 和 $L$）。如果可使用的资源限定为 $K$ 和 $L$，那末厂商的约束条件可用一个向量方程如下写出：

$$\begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} x_2 + \begin{bmatrix} 1 \\ 0 \end{bmatrix} s_1 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} s_2 = \begin{bmatrix} K_0 \\ L_0 \end{bmatrix} \quad (20.13)$$

纵向看(20.13)，即单独研究每个变量（及其系数向量）时，我

们可以把其中每个都看作代表厂商性质不同的**活动**。例如，第一种活动是生产第一种产品的活动，因而确定解值 $\bar{x}_1$ 的问题，实际上就是求出第一种活动的最佳水平的问题；同样，第二种产品的生产构成厂商的第二种活动；甚至可以将每个**松弛**变量也视为与某种独立活动有联系（这种活动“使某种资源不发挥作用”），虽然就其性质而言，与其说是活动不如说是不活动。

进行每种活动对厂商资源产生一定的影响。正如(20.13)中第一个向量说明的，第一种活动，每个单位消耗资本 $a_{11}$个单位，劳动 $a_{21}$个单位。与此类似，第二个向量表示第二种活动每个单位将耗费资本 $a_{12}$个单位，劳动 $a_{22}$个单位。我们把这些向量叫做活动向量。因此，它们表示所研究的活动水平增加一个单位所需的投入。注意，因为每个松弛变量只与一种特殊资源发生联系，所以其活动向量分别为单位向量 $e_1 \equiv \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 和 $e_2 \equiv \begin{bmatrix} 0 \\ 1 \end{bmatrix}$。

701

如果我们用 $A_1$ 和 $A_2$ 表示前两个活动向量，并把(20.13)右边的资源向量记成 $r$，向量方程(20.13)可更简单地写成：

$$A_1 x_1 + A_2 x_2 + e_1 s_1 + e_2 s_2 = r \qquad (20.13')$$

这个方程表明厂商的生产活动和松弛活动一起必恰好消耗全部可用资源。该方程也必是厂商生产线性规划中的约束条件。那么规划的目的就是选择上述四种活动的非负水平，使某个目标函数满足资源消耗约束条件(20.13′)而达到极大。用这样的观点考察线性规划问题时，它就变成一个活动分析的问题。

按照上述定义，一种活动向量仅表示该活动的**投入**需求。但向量中也可能包含**产出**情况。例如，我们描述第一种活动可以不

用向量 $\begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix}$，而用扩充向量 $\begin{bmatrix} 1 \\ -a_{11} \\ -a_{21} \end{bmatrix}$（加号表示产出，减号表示投入）来描述，活动；其中第一个元素表示单位产出水平，其余元素表示所需的投入。这里，这种标有投入需求形式的向量，更方便适用。

## 不变规模收益与固定投入比率

关于生产函数的两个假设，在活动分析结构中考察线性规划时，占有显著地位。这两个假设隐含在线性规划中。一个假设为不变规模收益（Constant Returns to Scale，简写为 CRTS）假设，另一个假设是固定投入比例假设。

CRTS 假设反映在下述事实中：活动向量 $A_1$ 和 $A_2$ 中的元素都是不变的。因为这种不变性，如果要求 $\begin{bmatrix} K \\ L \end{bmatrix}$ 的 $\begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix}$ 个单位生产 $x_1$ 产品的一个单位，那么要生产 $k$ 个单位 $x_1$ 就要求每种资源恰增加 $k$ 倍。第二种产品的情况也一样。事实上 CRTS 假设也适用于松弛活动。

固定投入比率的假设可从下述事实看出：每种产品仅由单个活动生产出来，对这种活动，如(20.13)所示，资本与劳动的比率由已知活动向量严格规定。这表明在我们的例子中劳动和资本的替代被完全排除了。当然，实际上线性规划并不是不可以代换的。如果生产产品 $x_1$ 的劳动和资本实际上可以组成（比方说）三种不同的比率，对所提到的产品 $x_1$ 组合成三种而不是一种独立活动，那末，就容易说明规划中的这种情况：

702

$$(A_1x_1 + A_1^*x_1^* + A_1^{**}x_1^{**}) + A_2x_2 + e_1s_1 + e_2s_2 = r \tag{20.14}$$

其中 $A_1 \neq A_1^* \neq A_1^{**}$ 为三个不同的活动向量，每个描述不同生产过程资本与劳动不同的比，且 $x_1$，$x_1^*$ 和 $x_1^{**}$ 分别表示在三种过程中生产第一种产品的数量。用这种方式，我们就能放宽投入比率的假设。但 CRTS 假设总是保持的。

第一种产品的三种生产过程恰好列在(20.14)中。但因为这些生产过程可以同时组合使用，所以生产第一种产品的方法实际上在三种以上。例如，如果厂商想要生产这种产品的总量为 12 个单位。这时或者令 $x_1 = 3$，$x_1^* = 5$ 和 $x_1^{**} = 4$，或者挑选组合(2, 9,1)或(6,0,6)，或(3.5, 4.5, 4)等等。为了生产这个总产出，这些可能组合中的每一组合都需要完全不同的资本与劳动比率。事实上，每一组合都可看成不同的生产方式。由此看来，一旦我们准许多种过程的生产进入线性规划，实际上固定投入比率已减弱到消失的地步。

但这是否意味着线性规划结构中所隐含的生产函数与古典生产理论中所出现的生产函数是一样的呢？答案是不完全一样。

## 生产函数：古典分析与线性规划

我们考察使用两种资源 $K$ 和 $L$ 的生产函数。为了便于与前面生产函数的讨论相联系，我们仍用符号 $Q$ 表示产出。

在古典分析中，我们假设生产函数

$$Q = f(K, L)$$

处处连续与可微。对某一具体产出水平，比方说 $Q = 1$，函数 $f$ 将产生一等产量曲线方程

$$1 = f(K, L)$$

其图形如图 20.1 所示，是一条光滑曲线。这条等产量曲线表示劳动和资本之间的连续替代性，且当我们持续替代时，替代比率也将连续地变化。关于所有其他可能的产出水平所作出的等量曲线，情况也一样。但古典生产函数可能有也可能没有 CRTS 的特性。

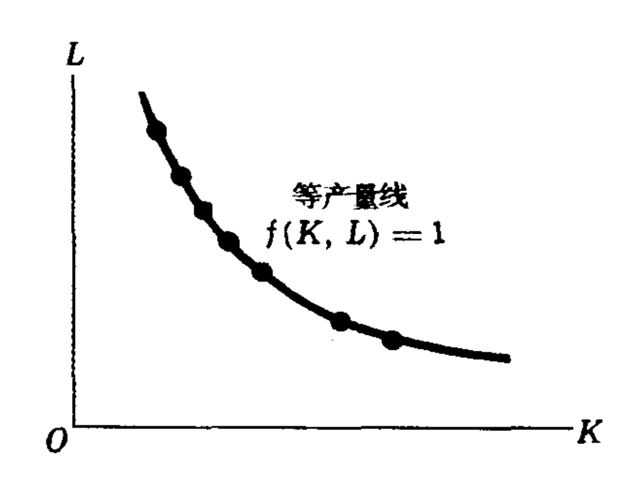



图 20.1

703 在线性规划或活动分析结构中，等产量曲线具有不同的表现形式。首先，我们考察单个生产过程的情况。假设第一种产品的常数活动向量为 $A_1 \equiv \begin{bmatrix} K_1 \\ L_1 \end{bmatrix}$，在这里将该产品看作仅有的产品。因为产出 $Q=1$ 仅由资本 $K_1$ 个单位的资本和劳动 $L_1$ 个单位生产出来，所以这个产出的等产量方程为：

$$1 = g(K_1, L_1)$$

由于 $K_1$ 和 $L_1$ 是两个常数，在图 20.2 中的 KL 平面内描出上述方程的图形是单个点 $A_1$，因而 $g(K_1, L_1)$ 仅指定义域中某具体点的生产函数值。当然，点 $A_1$ 只不过是上面所定义的活动向量的图形表示。由于线性规划中的 CRTS 假设，容易推出 $Q=2$ 的等产量曲线又是一单个点，该点为 $2A_1 \equiv 2\begin{bmatrix} K_1 \\ L_1 \end{bmatrix}$。而且因为所有

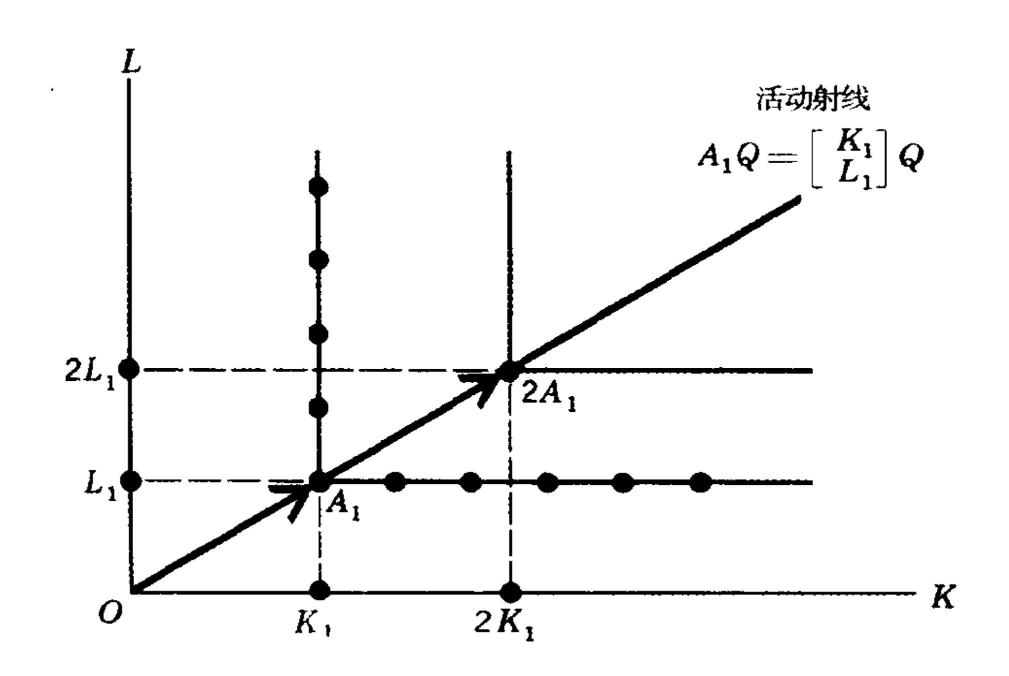



图 20.2

这样的等产量曲线点都是活动向量 $A$ 的非负倍数,所以它们都必然位于如图 20.2 所示的直线活动射线上。

对于单个点使用等产量曲线这个词似乎有点奇怪(一般这个词仅与曲线有联系)。实际上,把图 20.2 中的等产量曲线点重新解释成曲线也是可以的。我们知道资本保持在 $K_1$,而劳动增长超过 $L_1$ 时,根据固定投入比的假设,产出仍保持在 $Q=1$,由于没有同时增加资本,过多劳动是没有用处的,过多的资本也产生类似的结果。这些事实可用方程 $1=g(K_1,L_1+\delta)$ 和 $1=g(K_1+\delta,L_1)$ $(\delta\geqslant 0)$ 表示,在图形上它表示位于 $A$ 正东和正北方向的任意点得到产出 $Q=1$。因此,我们画出 $Q=2$ 的等产量曲线为一条 $L$ 形曲线,且在位于活动射线上的 $2A_1$ 处弯折,显然转折点(与曲线上其他任意点相比,该点对一种资源需求较少,对另一种资源的需求也不多)必是等产量曲线上最有效率的投入组合。因而受到转折点支配的其余各点可以不予考虑。 704

因此,单个生产过程的厂商必在活动射线

$$A_1Q = \begin{bmatrix} K_1 \\ L_1 \end{bmatrix} Q \ (Q \geqslant 0)$$

上运营。这里,$Q$ 是产出水平,也可解释为已知活动的运作水平。当我们考虑同种产品多过程的生产时,就会产生很多活动射线。如果(比方说)有四个活动向量:

$$A_1 \equiv \begin{bmatrix} K_1 \\ L_1 \end{bmatrix} \quad A_2 \equiv \begin{bmatrix} K_2 \\ L_2 \end{bmatrix} \quad A_3 \equiv \begin{bmatrix} K_3 \\ L_3 \end{bmatrix} \quad A_4 \equiv \begin{bmatrix} K_4 \\ L_4 \end{bmatrix} \tag{20.15}$$

705 其中各向量均表明能够生产单位产出的不同投入比率,则我们可画出四个活动向量,如图 20.3 所示。[①] 这些射线标为 $A_jQ_j(j=1,\cdots,4)$,其中 $Q_j$ 表示在第 $j$ 种生产过程所生产的产品数量。在多种生产过程的情况下,我们怎样求出 $Q(\equiv\sum_j Q_j)=1$ 的等产量曲线呢?首先,根据定义,四个点(每个点均表示在单位水平上"纯"过程的排他使用)应包含在所提到的等产量曲线上。但由于假设产品是可分的,也可以用多种不同的组合,例如

$$\frac{1}{2}A_1 + \frac{1}{2}A_2 \quad \frac{1}{3}A_2 + \frac{2}{3}A_3 \quad \frac{1}{5}A_3 + \frac{4}{5}A_4$$

来完成这些过程。只要组合的系数 $Q_j$ 加起来等于 1,我们就能得到产出 $Q=1$。在数学上,这表示为了生产单位产出,我们只须使用纯活动向量的**凸组合**;在图形上这表示关于**组合**过程的投入比必位于联结活动向量 $A_1,\cdots,A_4$ 的直线段上。

---

① 注意我们绘出活动射线 $A_1,\cdots,A_4$ 的方式:在生产一个单位产出需要较多的 $K$ 和 $L$ 的意义上,这些活动射线,没有哪一个居于"支配"地位,因而较多的 $K$ 和 $L$ 显然是没有用处的,所以也不会得到利用。

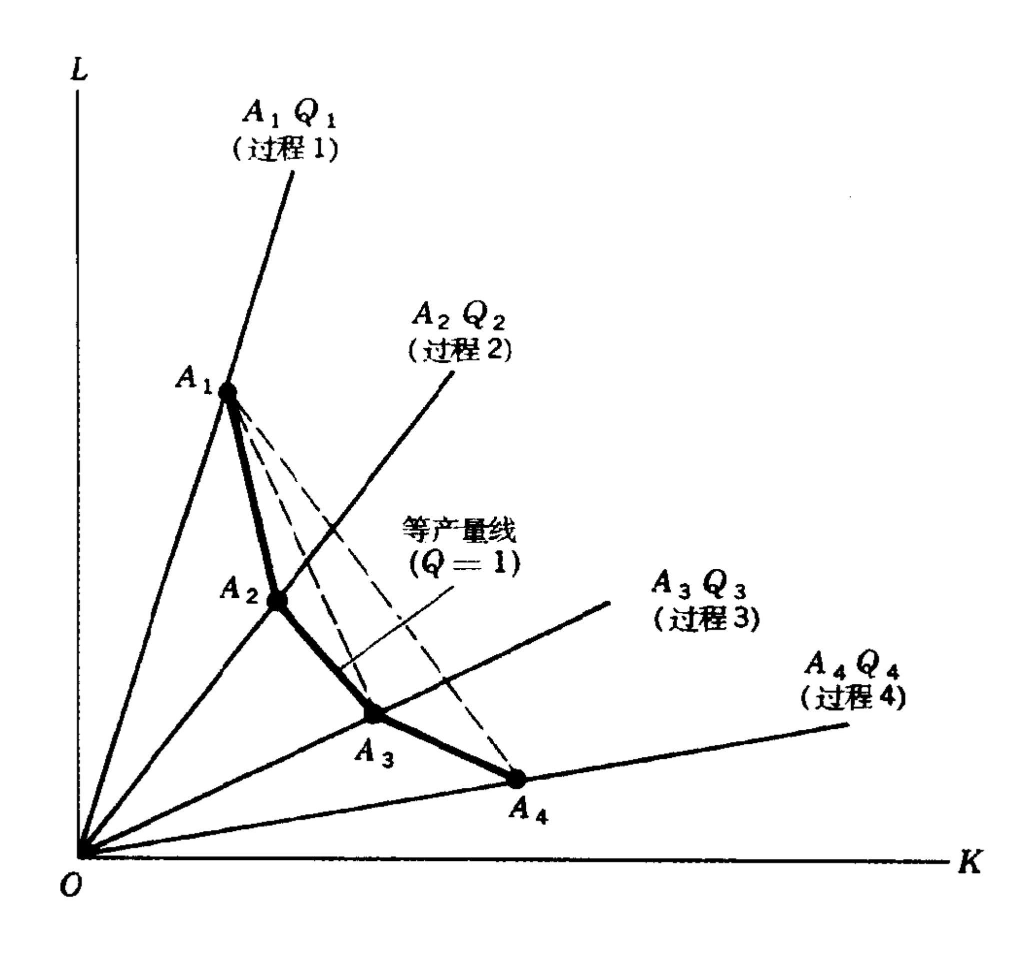



图 20.3

在这方面,我们应注意:虽然图 20.3 中的任意一对活动向量在理论上都可以是凸组合,但我们仅需考虑相邻向量(例如 $A_1A_2$ 和 $A_2A_3$)的凸组合。因为非相邻向量的组合无效,可以不予考虑。例如,(虚)线段 $A_1A_3$,因 $A_1A_3$ 上的各点(除端点 $A_1$ 和 $A_3$ 本身以外)位于 $A_1A_2$ 或 $A_2A_3$ 的东北方向,与 $A_1A_2$ 或 $A_2A_3$ 上各点相比较,含有较大的资本和劳动这两种投入需求,但所得到的产出与 $Q=1$ 相同。$A_1$ 和 $A_4$(彼此相隔最远的一对点)的凸组合情况更坏,因为可以看出虚线段 $A_1A_4$ 位于更远的地方,因而需要更多的投入。在这方面,显然线段 $A_1A_3$,$A_1A_4$ 和 $A_2A_4$(没有绘出)为 $A_1A_2$,$A_2A_3$,$A_3A_4$ 所支配,因而可以忽略。另一方面,所保留的三条线段上的点的确不互相支配。当我们从一个投入比转

到另一个投入比时,一种资源的增加伴随另一种资源的减少。因此,我们可取折直线 $A_1A_2A_3A_4$ 作为 $Q=1$ 的等产量曲线。

这条等产量曲线与图 20.1 所示的古典光滑等产量曲线显然不同。虽然等量线 $A_1A_2A_3A_4$ 较图 20.2 中单过程情况下的等产量曲线更加接近于古典的形式,但它不光滑。替代率在某一范围内往往保持不变,但也有突然的跳跃。此外,只有厂商使用的纯过程仅有有限多个,这个特征就能保持。因为过程数目增长为任意 706 有限个,等量线上只增加直线段的数目,但并不能消除其上的角点。因此,就此而论,古典理论的光滑等产量曲线显然必须建立在这样的假设基础之上:厂商可采用的生产过程的数量是**无限**的。

另一产出水平的等量曲线可用同样的方法作出。例如,对 $Q=2$,我们可简单地把每个向量 $A_j$ 延长到 $2A_j$(该点的纵坐标和横坐标加倍),然后将产生的新向量点的每相邻两点相连。根据线性规划中 CRTS 的假设,这个弯折的新等产量曲线上的所有点将同产出 $Q=2$ 有关。在图 20.4 中已绘出四条这样的等量曲线,但同古典方法比较,其中还有其它差别。因为虽然古典等产量曲线可能具有 CRTS 特性(相差均匀间隔),但不一定必然是这样。可是,在线性规划中总假定具有 CRTS 的特性。

注意,虽然也允许把位于 $A_1$ 正北方的点和位于 $A_4$ 正东方的点加到第一条等产量曲线($Q=1$)上,对其他等产量曲线也可类似考虑,但我们不这样做。原因很简单,这样的点是被支配的点。因此,所有等产量曲线将位于以两根最远的活动射线 $A_1Q_1$ 和 $A_4Q_4$ 为界的锥形面之内。这样的锥形面组成的凸集,叫做**凸锥面**。表示两个向量(这里是 $A_1$ 和 $A_4$)的所有非负线性组合的集合。即,如果我们取遍两个向量所有可能的非负倍数,且由平行四

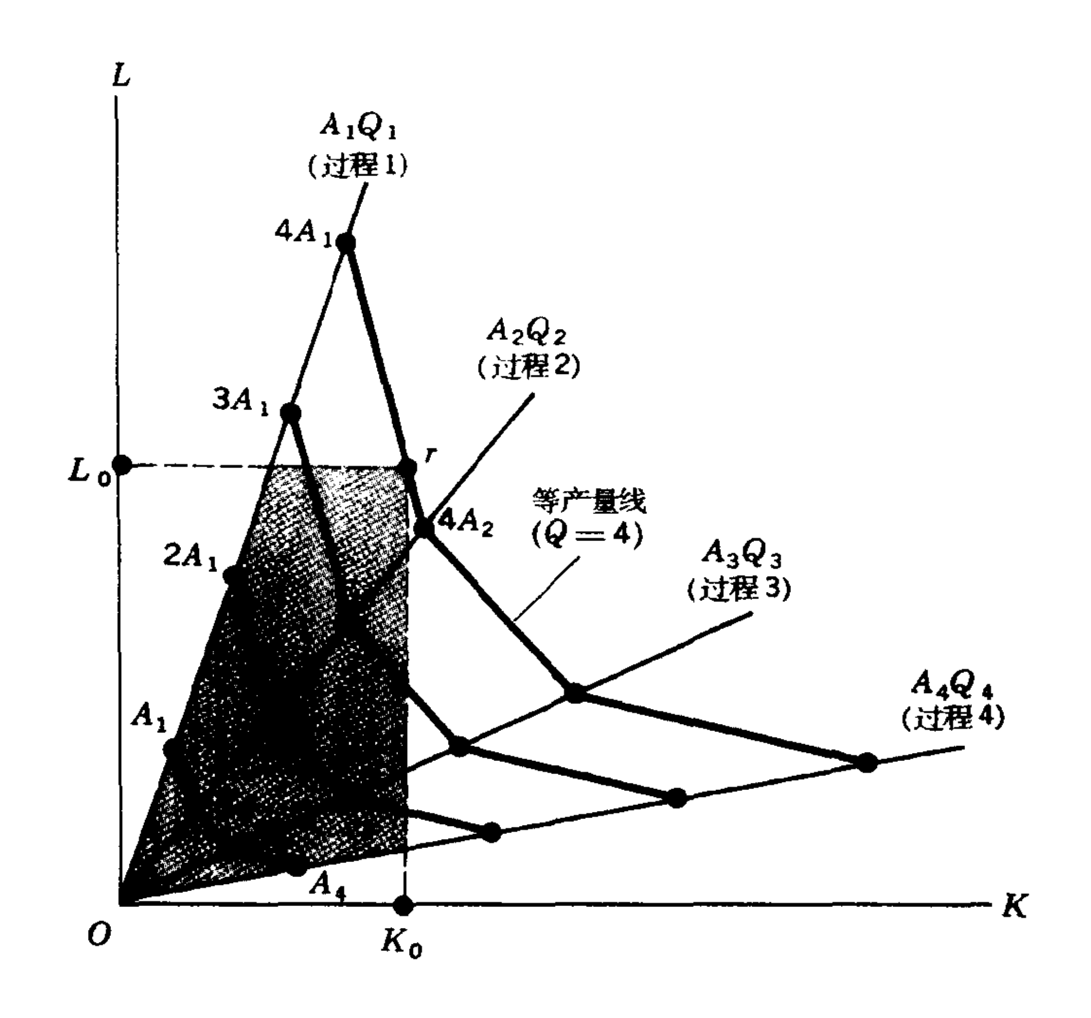



图　20.4

边形法则作出所有可能的和向量，这些和向量将填满这锥形面。

## 最优化



现在我们根据约束条件的活动解释，列出简单的线性规划的表达式。假设在图 20.4 中所描述的厂商通过多过程生产单一产品，在其一定资源 $K_0$ 和 $L_0$ 的范围内，设法使产出 $Q$ 极大化，那么有关的线性规划将是

$$\begin{aligned} \text{Max.} \quad & Q = Q_1 + Q_2 + Q_3 + Q_4 \\ \text{s.t.} \quad & A_1Q_1 + A_2Q_2 + A_3Q_3 + A_4Q_4 \leqslant r \qquad (20.16) \\ \text{且} \quad & Q_j \geqslant 0 \ (j = 1, \cdots, 4) \end{aligned}$$

其中所有 $Q$ 符号都是标量，符号 $A_j$ 表示(20.15)中所定义的活动向量，$r \equiv \begin{bmatrix} K_0 \\ L_0 \end{bmatrix}$ 表示现有资源。鉴于现在的目的，我们忽略松弛性活动。

尽管选择变量有四个，但我们仍能用图解法解出这个规划。因为我们沿纵向读出约束条件，且每个活动向量是 2 维向量，在图 20.4 的 $KL$ 坐标平面(投入空间)中绘出约束条件中每项 $A_jQ_j$ 的图形为一活动射线。因此，将资源而不是选择变量标在坐标轴上时，四维的问题就简化为二维的问题了。

因为已把等产量曲线(它包括经济上有效的投入组合)都限定在两条最远的活动线所确定的凸锥面以内，所以厂商确实不必在锥面外去寻求最优解。而资源的限制进一步缩小了选择的范围。如果在图中描出 $K_0$ 和 $L_0$，那么矩形区域 $OK_0r\,L_0$，将确定供厂商选择的所有可行投入组合的集合；去掉位于凸锥面以外的矩形，最后我们得到的阴影区域(矩形与锥面的交)就是最大的选择范围。

使产出 $Q$ 极大化意味着上升到尽可能高的等产量曲线。为了满足资源约束条件，我们看出图 20.4 中所得到的极大产出是 $Q=4$，且在点 $r$ 处要求充分利用现有的资源 $K_0$ 和 $L_0$。因为点 $r$ 位于联结 $4A_1$ 和 $4A_2$ 的直线上，显然，在最优解的情况下，只有选定头两种生产过程，即我们必须令 $\bar{Q}_3=\bar{Q}_4=0$。这个事实，加上 $K_0$ 和 $L_0$ 被充分利用的事实，意味着 $A_1Q_1+A_2Q_2=r$，或者：

$$\begin{bmatrix} K_1 \\ L_1 \end{bmatrix} Q_1 + \begin{bmatrix} K_2 \\ L_2 \end{bmatrix} Q_2 = \begin{bmatrix} K_0 \\ L_0 \end{bmatrix} \tag{20.17}$$

由上面含两个方程的线性方程组，容易求出：

$$\bar{Q}_1 = \frac{K_0L_2 - K_2L_0}{K_1L_2 - K_2L_1} \quad 和 \quad \bar{Q}_2 = \frac{K_1L_0 - K_0L_1}{K_1L_2 - K_2L_1}$$

也可用其它方法来解决这个问题。因为 $r$ 是 $4A_1$ 和 $4A_2$ 的凸组 708 合,所以在区间[0,1]中必存在唯一的标量 $\theta(4A_1)+(1-\theta)(4A_2)=r$,即:

$$4\theta\begin{bmatrix}K_1\\L_1\end{bmatrix}+4(1-\theta)\begin{bmatrix}K_2\\L_2\end{bmatrix}=\begin{bmatrix}K_0\\L_0\end{bmatrix} \tag{20.18}$$

由此得出

$$\bar{\theta}=\frac{K_0-4K_2}{4(K_1-K_2)} \quad \left[也=\frac{L_0-4L_2}{4(L_1-L_2)}\right]$$

那么,使(20.17)和(20.18)相等,我们也可把解值表示为:

$$\bar{Q}_1 = 4\bar{\theta} = \frac{K_0-4K_2}{K_1-K_2} \quad 和 \quad \bar{Q}_2 = 4(1-\bar{\theta}) = 4 - \frac{K_0-4K_2}{K_1-K_2}$$

当然,上述问题也可用一般的单纯形法来解。或者我们也可以先写出线性规划的对偶规划,然后解这个对偶规划,从而得到原规划的解。在这两种情况下,最优解的结果是一样的。特别是,因为有两个约束条件,我们知道仅有两个选择变量 $Q_j$ 在最优解中必须为非零值。这恰是我们刚才由问题的活动分析形式所得到的结果。

## 练习 20.3

---

1 假定下面 $\begin{bmatrix}K\\L\end{bmatrix}$ 的三种组合中任意一种都可生产一单位产出:

$$A_1=\begin{bmatrix}2\\5\end{bmatrix} \quad A_2=\begin{bmatrix}3\\2\end{bmatrix} \quad A_3=\begin{bmatrix}5\\1\end{bmatrix}$$

(*a*) 作图表示向量 $A_1, A_2$ 和 $A_3$。

(*b*) 作出 $Q=1$ 的等产量曲线。该等产量曲线含有多少个转折点?

(*c*) 画出活动射线 $A_1Q_1, A_2Q_2$ 和 $A_3Q_3$。

(*d*) 根据 CRTS 的假定,作出 $Q=2$ 的等产量曲线。

2 在上一问题中,如果还知道第四个活动 $\begin{bmatrix}3\\4\end{bmatrix}$ 也能生产一单位产出,$Q=1$的等产量曲线上能包含它吗?为什么能?为什么不能?

3 在练习 20.3-1 中,$Q=1$ 的等产量曲线包含下面三个点吗?

$$(a)\ (\frac{5}{2},\frac{7}{2}) \quad (b)\ (\frac{8}{3},3) \quad (c)\ (\frac{7}{2},\frac{7}{4})$$

4 假设厂商在生产某产品时,有四种活动,利用三种资源

(*a*) 其活动射线位于 3 维空间还是 4 维空间呢?

(*b*) 这些射线仍能确定一凸锥体吗?

709 (*c*) 这个锥体更可能像冰淇淋锥体还是切断的棱锥?为什么?

5 在图 20.4 的问题中,如果资源约束变为:

$$(a)\ \begin{bmatrix}K_0\\L_0\end{bmatrix}=\begin{bmatrix}3K_4\\3L_4\end{bmatrix} \quad (b)\ \begin{bmatrix}K_0\\L_0\end{bmatrix}=\begin{bmatrix}K_2+3K_3\\L_2+3L_3\end{bmatrix}$$

其解会如何改变?

6 在图 20.4 的问题中,如果点 $r$ 移动到 $4A_1$ 正北的某个位置,新的最优解等于什么?其类型与图 20.4 中所示的解有什么基本不同?

7 某厂商使用单生产过程,其生产函数具有如图 20.2 所说明的 CRTS 和固定投入比的两个特性,证明该厂商的生产函数可用方程 $Q=\min\left\{\frac{K}{K_1},\frac{L}{L_1}\right\}$ 表示。

---

## 20.4 活动分析:宏观水平

前一节所介绍的大部分内容也可应用到宏观水平的分析。在厂商水平,每一活动都与独立的生产过程有关。在国家水平上,我们把每种活动视为表示一个产业部门。相应地,活动进行的水平

现在将是确定整个产业部门(单个)产品的产出水平。如果假设每个产业部门都具有规模收益不变和固定投入比的特征,而且每个产业部门投入—产出关系可用唯一的一条活动射线来概括,那么,若在射线上一点到原点的距离为另一点到原点的距离的两倍,则该点表示的产出也为其两倍。

由于CRTS和固定投入比在投入—产出模型中是标准的假设,似乎现在我们能够使用活动分析或线性规划解释投入—产出分析。而情况也的确是这样。

## 投入—产出分析和线性规划

进行每种活动(产业部门)所需要的投入有两种类型:**原始**投入和**中间**投入。前者不是产业部门的产出,而后者本身则是其他产业部门的产出。假设产业部门总数为 $n$ 个,且每个产业部门生产不同的商品。如果使用5.7节中相同的符号表示投入系数,我们会发现下述三种类型的向量很重要:

$$\begin{bmatrix} 1 \\ -a_{0j} \\ -a_{1j} \\ -a_{2j} \\ \vdots \\ -a_{nj} \end{bmatrix} \quad \begin{bmatrix} a_{0j} \\ a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix} \quad \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix} \tag{20.19}$$

第一类向量给出了第 $j$ 个活动(产业部门)的投入—产出关系 710 的完整描述:其中第一个元素(1)表示 $j$ 种商品的单位产出;第二个元素表明对这种商品的原始投入需求;其余元素描述中间投入

需求。第二类向量中去掉了单位元素，且所有的 $a_{ij}$ 的符号相反，它仅集中描述该产业部门的投入方面。第三个向量甚至进一步缩小了描述的范围，仅列出中间投入。对(20.19)中的第一种形式，我们通常给予它一个专门名称：活动向量。但是正如我们在上一节中的做法一样，我们也反过来用这个名称表示它的修正形式，对此，我们后面要加以介绍。

现在，我们考察(20.19)中的第三种向量形式。因在经济中有 $n$ 个产业部门，我们可写出 $n$ 个这样的向量。把它们集中在一起，就构成一个大家熟悉的 $n \times n$ 的矩阵：

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

给定此矩阵，以及下面的产出向量和最终需求向量：

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad d = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$$

则我们的任务与前面讨论的开放投入—产出模型一样，就是要求向量 $x$，使

$$(I - A)x = d \tag{20.20}$$

假定 $(I - A)$ 为非奇异矩阵，其解就是

$$\bar{x} = (I - A)^{-1}d \tag{20.21}$$

显然，问题的这种表述方式对线性规划最优化的类型不会有

任何启示，因为它并没有优选的目标函数，而且虽然方程(20.20)对每个产业部门的产出水平具有约束条件的性质(每个产业部门应生产足够的产出以满足总需求)，却根本没有包含不等式。

同一个投入—产出问题可从不同角度来考察。首先容易看出，为了保证满足总需求，只要每个产业部门的产出不少于(而不是等于)对它的总需求即可。因此，把(20.20)变为下述不等式是可以的：

$$(I - A)x \geqslant d$$

但为了防止不必要的过剩产出(即防止≥号中>部分"失去控制")，我们应对这个不等式附上某种极小化的要求。例如，假定劳动是仅有的原始投入，我们可设法使生产上述产出所要求的总劳动投入为极小，即我们尽力使：

711

$$L = \sum_{j=1}^{n} a_{0j}x_j = [a_{01}\ a_{02}\ \cdots\ a_{0n}]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A'_0 x$$

极小化。其中 $L$ 表示所需要的总劳动，$A'_0$ 表示劳动—投入系数的行向量。此外，因产出水平 $x_j$ 决不为负，施加限制条件 $x \geqslant 0$ 也是合理的。为此，投入—产出模型(20.20)可以用等价的数学形式重新表述如下：

$$\begin{aligned} &\text{Min.} \quad L = A'_0 x \qquad\qquad (20.22) \\ &\text{s.t.} \quad (I - A)x \geqslant d \\ &\text{且} \quad x \geqslant 0 \end{aligned}$$

读者会认出，这只不过是标准的线性规划而已。

## 解 法

线性规划(20.22)有两种解法。一种解法是按水平方向读出 $n$ 个约束条件,而另一种解法则是按纵向读出约束条件。

按水平方向读出约束条件时,$n$ 个约束条件产生 $n$ 个闭半空间,与非负限制一起,在非负的正交分划体中确定可行区域(凸集)。另一方面,目标函数生成一族等劳动超平面。求出最优解就是在可行区域内选定这样一点,使该点位于 $L$ 值最小的等劳动超平面上。但该点($\bar{x}$)必然是(20.21)中所给出的一点。因为如果劳动是生产每种商品必不可少的投入,所以,劳动需求最小的产出向量必然是这样的向量:它不含超过总需求的过剩产出,即最优产出向量必然是 $\bar{x}=(I-A)^{-1}d$,这是一般的投入—产出模型的解。注意,在这种情况下,所有约束条件都是满足的。产生这个结果的原因在于,本问题的选择变量的个数与约束条件的个数相等。

现在我们按纵向读出约束条件。因为矩阵($I-A$)由下列 $n$ 个向量

$$\begin{bmatrix}1-a_{11}\\-a_{21}\\\vdots\\-a_{n1}\end{bmatrix}\quad\begin{bmatrix}-a_{12}\\1-a_{22}\\\vdots\\-a_{n2}\end{bmatrix}\cdots\begin{bmatrix}-a_{1n}\\-a_{2n}\\\vdots\\1-a_{nn}\end{bmatrix}$$

所组成,所以,我们可以在 $n$—空间将其绘成活动向量。准确地说,我们应在最终需求空间(第 $j$ 个轴表示对 $x_j$ 的最终需求)中将其绘出——就像(20.13)中的活动向量应在投入空间(KL 空间)绘出一样。然后,通过一组非负系数 $x_1,\cdots,x_n$ 将这些向量线性组合起来,所有这些非负组合的集合呈凸多面锥的形式——它是

712

$n$ 维的，与图 20.4 中遇到的凸锥面相类似。在图 20.5 中对 $n=3$ 的情况给出对这种锥面的直观图像。如果三个箭号确定三个向量，那么三个向量中任意两个的所有非负组合(由平行四边形确定)的集合组成了像 $F_1$ 和 $F_2$ 这样平直的三角形曲面，它们是 2 维的，且类似于图 20.4 中的凸锥面。如果我们进一步作出 $F_1$ 和 $F_2$ 上点的所有可能的非负组合，我们也能填满由面 $F_1$ 和 $F_2$ 和 $F_3$(藏在背后)所围成的空间。于是，这三个已知向量非负组合的实体是那些位于图 20.5 中形似角锥体的界面或位于其内部的点的集合这个锥体像 $F_1$ 和 $F_2$ 一样，被假定为始终向外延伸。这个锥体构成了一个凸多面锥体(一个凸集)，由 $F_1$、$F_2$ 和 $F_3$ 表示的面称为锥体的面或小平面。

尽管在图 20.5 中所绘出的直观凸多面锥体完全在非负的正

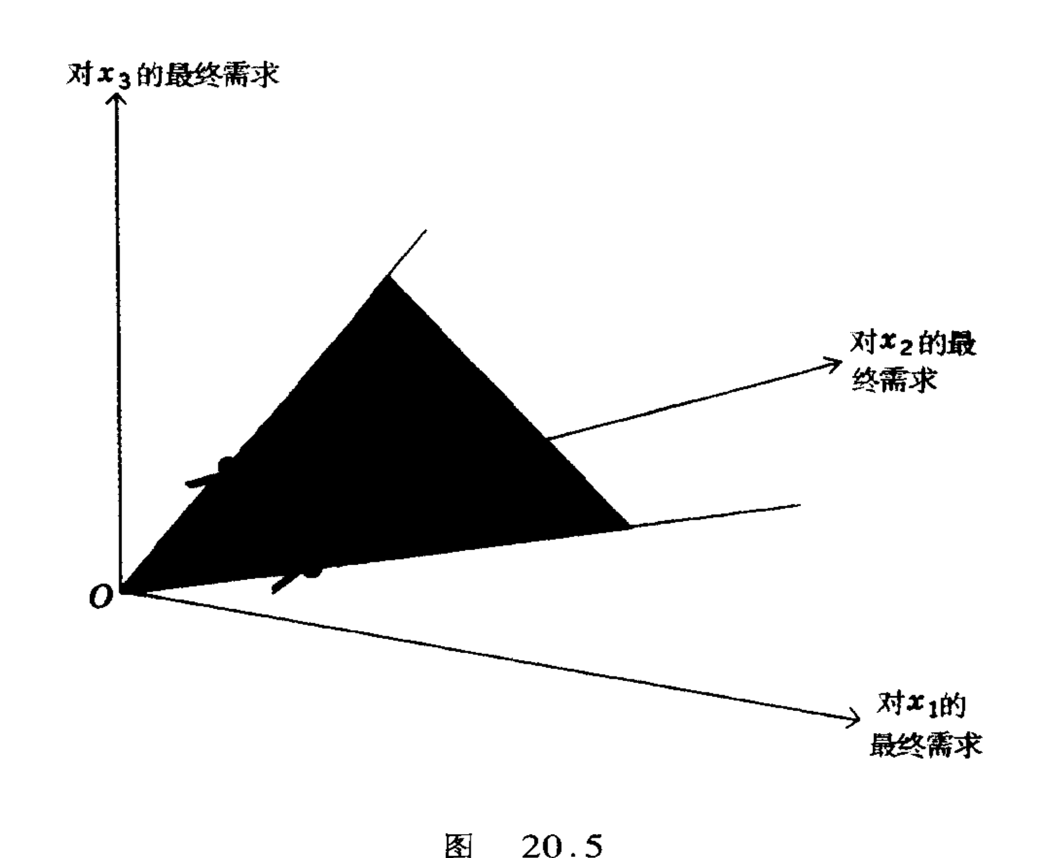



图 20.5

713 交分划体中，但在投入—产出线性规划中，情况则不是这样。在三个产业部门的模型中，约束条件取下列形式：

$$\begin{bmatrix} 1-a_{11} \\ -a_{21} \\ -a_{31} \end{bmatrix} x_1 + \begin{bmatrix} -a_{12} \\ 1-a_{22} \\ -a_{32} \end{bmatrix} x_2 + \begin{bmatrix} -a_{13} \\ -a_{23} \\ 1-a_{33} \end{bmatrix} x_3 \geqslant \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \tag{20.23}$$

因为所有系数 $a_{ij}$ 或者为正，或者零，所以三个活动向量的元素（在非零的情况下）加上符号应为：

| 向量 1 | 向量 2 | 向量 3 |
|---|---|---|
| $\begin{bmatrix} + \\ - \\ - \end{bmatrix}$ | $\begin{bmatrix} - \\ + \\ - \end{bmatrix}$ | $\begin{bmatrix} - \\ - \\ + \end{bmatrix}$ |

由于向量中出现负元素，活动向量必须位于非负正交分割体的外面。图 20.6$a$ 中给出一个 3 维空间的俯视图。例如，在这个图中看到射线 1 在 $x_1$ 方向为正，但在 $x_2$ 方向为负；实际上在 $x_3$ 方向上也是负，即它远离我们的视点延伸。对射线 2 也可作类似解释。在图 20.6$b$ 的 3 维空间中绘出其图形时，这些射线必然会产生一个包含非负正交分割体的凸多面锥体。这与图 20.5 中的情况恰好相反。

714 我们现在已把(20.23)左边的三个向量的非负组合解释为对应于凸多面锥体中的点。我们如何解释剩下的表达式$\geqslant d$ 呢？坐标为$(d_1, d_2, d_3)$的向量点 $d$ 直接位于图 20.7 中基平面点$(d_1, d_2, 0)$的正方上。当第 $j$ 种商品的需求不少于具体数量 $d_j(j=1,2,3)$的条件满足时，表示$\geqslant d$ 确定最终需求空间一个子集。如果我们

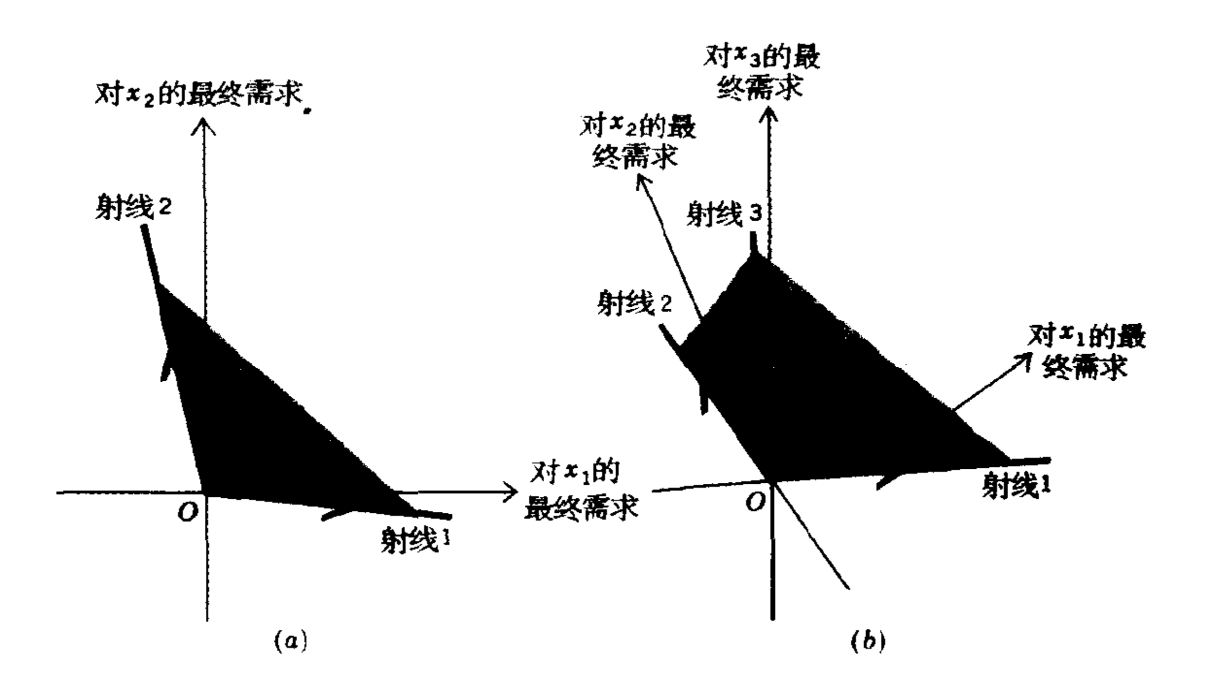



图 20.6

取点 $d$ 作为新原点，平行于原坐标轴绘出三根新轴（见箭头），那么提到的子集对于 $d$ 点的关系与非负正交分割体对原点是一样的。由于它看起来像房子上层的房间，所以，我们称其为阁楼 $d$。

从整体上考察(20.23)，要求我们把注意力集中在凸多面锥体中位于阁楼 $d$ 的那些点上，但因为阁楼 $d$ 是非负正交分划体的子集，因此，是锥体的子集，这表明需考虑的只有阁楼 $d$。

当 $L$ 仍在阁楼 $d$ 上，线性规划的目标就是使 $L$ 极小化，即从等劳动平面（或更准确地说等劳动平面部分）族中选取与阁楼 $d$ 有一公共点的最底平面。因这样的平面部分(plane segment)都平行，其典型形状似图 20.7 中由阴影构成的三角形（见练习 19.4 第 4 题），所以最优的一个平面部分是在点 $d$ 与阁楼 $d$ 接触的那个平面部分。因此约束条件(20.23)在最优的情况下，将是严格等式，这又使我们回到公式(20.20)，得到与(20.21)相同的解。

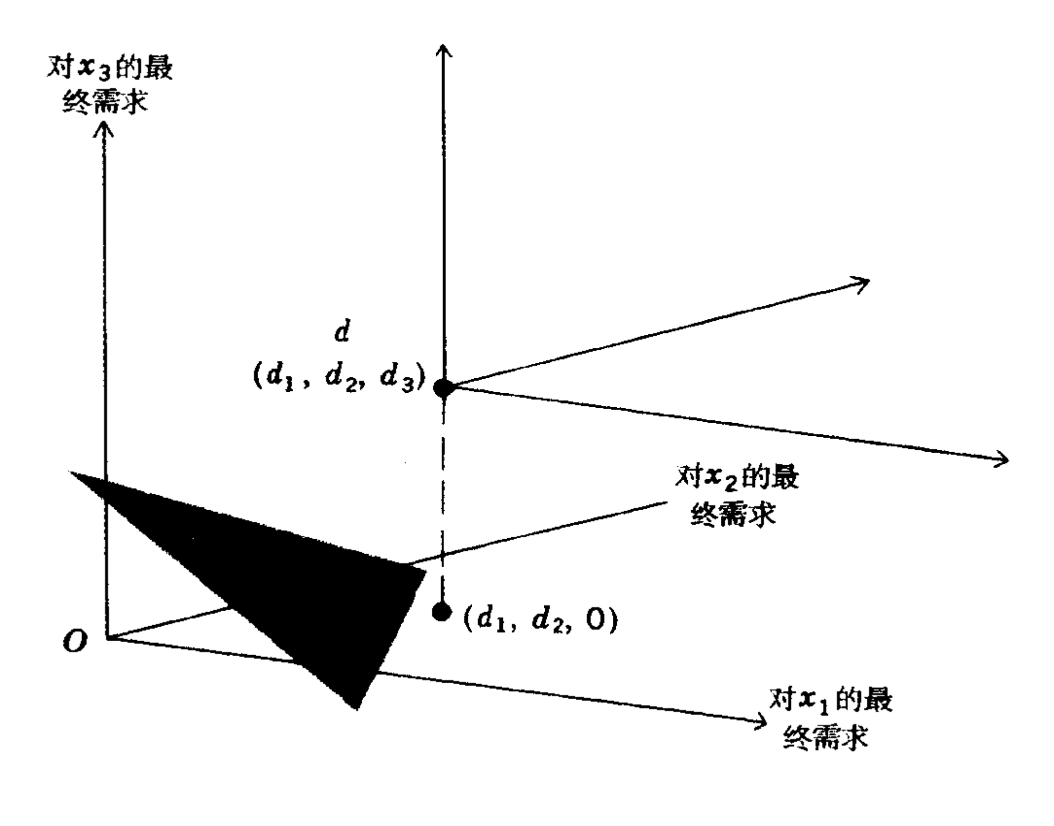



图 20.7

## 715 练习 20.4

1 文中所讨论的投入—产出模型中,劳动利用没有加以限制。如果现有的劳动仅有限的数量 $L_0$,对模型应施加什么额外的约束条件?分别用符号与向量符号写出来。

2 原点是一锥体元素(称为锥体的顶点),从活动分析的观点的解释为什么应是这种情况?

3 已知三个向量 $\begin{bmatrix}1\\1\end{bmatrix}$, $\begin{bmatrix}2\\-1\end{bmatrix}$ 和 $\begin{bmatrix}3\\1\end{bmatrix}$,试绘出由这些向量的非负组合所生成的锥面。

4 ($a$)对图 20.7 中所示的形似三角形的等劳动平面部分给出解释(提示:回顾对图 20.3 中等产量曲线结构的讨论)。

($b$)等劳动平面部分为什么都平行?

# 第 21 章　非线性规划

前两章所讨论的线性规划，就实际价值而言，显然是对古典的 716 最优化结构的一个很大的改进，因为约束作为不等式出现，现在已成为线性规划问题的组成部分。因此，在线性规划问题中我们也能明确地引入非负限制。但是，把目标函数和约束条件限定为线性模式，有时候会变成一种严重的缺陷。因此，作为进一步的改进，我们需要一种既可以处理非线性目标函数，也能处理非线性不等式约束的最优化结构。这样的结构可从非线性规划中找到。

## 21.1　非线性规划的性质

非线性规划的极大化问题具有下述一般形式：

$$
\begin{aligned}
\text{Max.}\quad & \pi = f(x_1, x_2, \cdots, x_n) \\
\text{s.t.}\quad & g^1(x_1, x_2, \cdots, x_n) \leqslant r_1 \\
& g^2(x_1, x_2, \cdots, x_n) \leqslant r_2 \\
& \cdots\cdots\cdots\cdots\cdots\cdots \\
& g^m(x_1, x_2, \cdots, x_n) \leqslant r_m \\
\text{且}\quad & x_j \geqslant 0 \quad (j = 1, 2, \cdots, n)
\end{aligned}
\tag{21.1}
$$

用不带下标的符号 $x$ 表示选择变量的有序 $n$ 元素组时，可以把 717 (21.1)更简洁地写成：

$$\text{Max.}\quad \pi = f(x)$$

$$\text{s.t.} \quad g^i(x) \leqslant r_i \qquad (i=1,2,\cdots,m)$$

$$\text{且} \quad x \geqslant 0 \tag{21.1'}$$

同样,极小化问题也可写成如下形式:

$$\begin{aligned}
\text{Min.} \quad & C = f(x_1, x_2, \cdots, x_n) \\
\text{s.t.} \quad & g^1(x_1, x_2, \cdots, x_n) \geqslant r_1 \\
& g^2(x_1, x_2, \cdots, x_n) \geqslant r_2 \\
& \cdots\cdots\cdots\cdots\cdots\cdots \\
& g^m(x_1, x_2, \cdots, x_n) \geqslant r_m \\
\text{且} \quad & x_j \geqslant 0 \ (j=1,2,\cdots,n)
\end{aligned} \tag{21.2}$$

或更简洁地写成

$$\begin{aligned}
\text{Min.} \quad & C = f(x) \\
\text{s.t.} \quad & g^i(x) \geqslant r_i (i=1,2,\cdots,m) \\
\text{且} \quad & x \geqslant 0
\end{aligned} \tag{21.2'}$$

根据上述内容,显然可以知道,非线性规划也同线性规划一样,由三部分组成:目标函数,一组($m$ 个)约束条件和 $n$ 个选择变量的一组非负限制。同在线性规划中一样,$m$ 可大于,等于或小于 $n$。非线性规划问题中的所有函数 $f(x)$和 $g^i(x)$都假定为可微。还要注意,为了与前面讨论的线性规划一致,我们在极大化问题中仅采用$\leqslant$型的不等式约束,在极小化的问题中仅采用$\geqslant$型的不等式的约束。当出现类型相反的不等式约束时,在不等式两边通乘 $-1$,便可以很容易地将其变成上述类型的约束。

以(21.1)和(21.2)给出的公式,显然是本书迄今为止所见到的最一般形式的最优化问题。因此,它们包含了前面所讨论的各种最优化问题。显然,线性规划是它的特殊情况。如果我们仅考察目标函数,就得到自由极值问题。最后,如果把用严格不等式形

式表示的约束条件部分加入目标函数，且 $m<n$，又得到古典约束最优化问题。[①]

## 经济学中的非线性问题

718

在经济学中，非线性的问题可能以各种形式产生。在线性规划的生产问题中，我们假设每种产品的单位总利润是常数。但它完全可能是产出水平的递减函数。这既可能是因为大量的产品会使市场价格(平均收益)跌落，也可能是因为增加生产使产品的平均可变成本上升。如果是这样的话，线性目标函数$\pi=c_1x+\cdots+c_nx_n$ 必然为非线性形式，例如，$\pi=c_1(x_1)x_1+\cdots+c_n(x_n)x_n$ 所代替，其中 $c_j(x_j)$表示变量 $x_j$ 的递减函数。

类似地，在约束条件中，生产产品 $j$ 对第 $i$ 种资源的投入需求，可能随产品 $j$ 的产出水平而递减。例如，可以想象，后期的产品生产比开始时速度要快得多，因而每个后继的单位产出所花费的机时要少得多。当然，这就破坏了线性规划中系数 $a_{ij}$为常数的假定，也可能系数 $a_{ij}$不仅依赖于产品 $j$ 的产出水平，而且还依赖于另一种产品 $k$ 的产出水平。因而，在约束条件部分就会出现包含两个变量 $x_j$ 和 $x_k$ 乘积的项。于是，又将失去线性性质。

当上面所说明的经济情况在我们所研究的问题中出现时，以非线性的表达式表示比用线性方式表示总是更恰当一些。遗憾的

---

① 实际上，非线性规划和古典最优化之间的联系，比刚才简单说明的更为密切。如果想要知道这种联系，我们可采取下述两个步骤把非线性规划问题变成古典最优化问题：(1)引进 $m$ 个虚拟变量，把不等式约束变成等式约束。(2)把 $n$ 个选择变量和 $m$ 个虚拟变量中的每个变量看作一个新人工变量的平方，以保证非负性。(见练习 21.1—6)

是，线性规划的很多方便的特性，在那时就不适用了。可用一些简单的非线性规划例子来说明这个事实，这些例子都可用图解法求解。

## 图解法

这里我们介绍三个具体例子，每个例子都因具有区别非线性规划与线性规划之间的某些特性而受到人们的注意。

**例 1**　Min.　$C=(x_1-4)^2+(x_2-4)^2$

s.t.　$2x_1+3x_2\geqslant 6$

$-3x_1-2x_2\geqslant -12$

和　$x_1, x_2\geqslant 0$

因为这个问题的约束条件是线性的，可行区域的形状确实与线性规划有根本的不同。如图 21.1$a$ 中的阴影区域所示，由第一个约

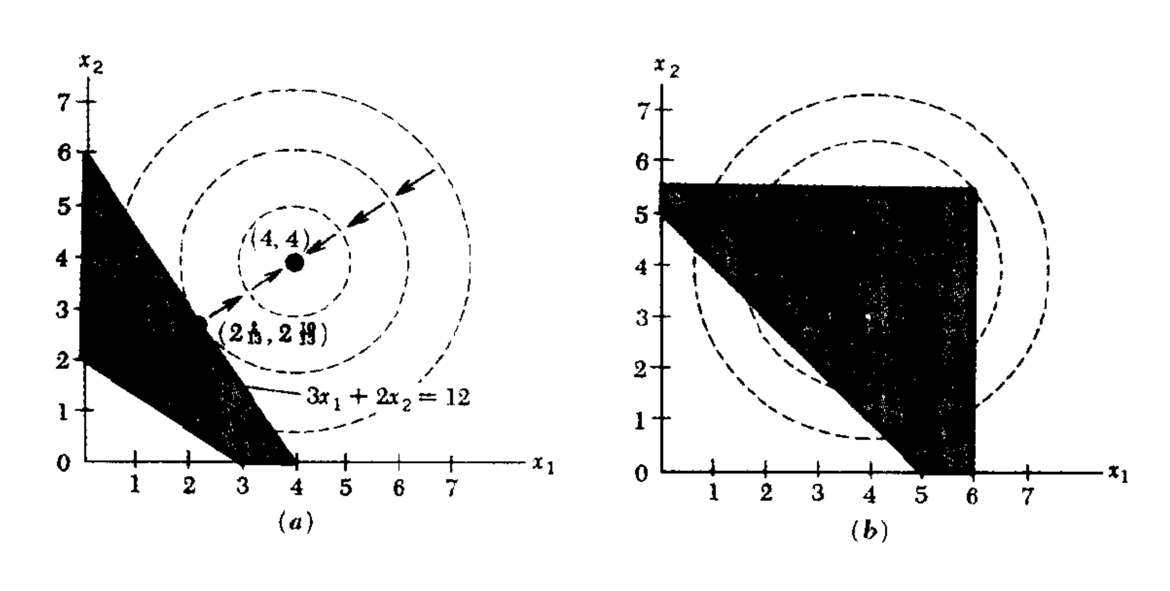



图　21.1

719 束条件可得到可行区域的西南边界，由第二个约束条件可得到其东北边界。因为目标函数是非线性的，它不能产生一族平行的等值直线，得到的是一族同心圆，其圆心在点(4,4)，每个相继变小的

圆都与较小的 $C$ 值相联系。当然，在自由极值的问题中，我们可选点$(x_1, x_2)=(4,4)$得到极小值 $C=0$。但要限制在阴影区域上，只好选点$(\bar{x}_1,\bar{x}_2)=(2\frac{2}{13}, 2\frac{10}{13})$，该点是东北边界与某一圆的切点。则 $C$ 的极小值为$\bar{C}=(2\frac{2}{13}-4)^2+(2\frac{10}{13}-4)^2=4\frac{12}{13}$。

虽然 $\bar{x}_1$和 $\bar{x}_2$的精确值是以几何上的切点为基础的，但也可以通过代数方法求出。首先，因最优解位于东北边界上，它必满足方程

$$3x_1+2x_2=12 \qquad \text{[由第二个约束条件]}$$

其次，与边界相切的圆在切点必与此边界有相等的斜率，即 $-3/2$。又因该圆的斜率是[对方程 $F(x_1,x_2)=(x_1-4)^2+(x_2-4)^2-C=0$ 使用隐函数法则]：

$$\frac{dx_2}{dx_1}=-\frac{\partial F/\partial x_1}{\partial F/\partial x_2}=-\frac{2(x_1-4)}{2(x_2-4)}=-\frac{x_1-4}{x_2-4}$$

可见，令其等于 $-3/2$，我们得到另一方程

$$2x_1-3x_2=-4$$

联立解这两个方程，便得到最优值$(\bar{x}_1, \bar{x}_2)=(2\frac{2}{13}, 2\frac{10}{13})$。

这里要注意，正如我们在线性规划中所预期的那样，最优解不在可行区域的极点上。因而，我们看到只有一个而不是两个约束条件被满足。还要注意，当沿东北方向接近点(4,4)时，最初使 $C$ 递减。然而继续沿同一方向移动，越过这一点时，便会得到较大的 $C$ 值。这样，再也没有理由像在线性规划中那样，沿某个单一的特定方向尽可能远地移动等值曲线得到极值解了。 720

**例 2** Min. $C=(x_1-4)^2+(x_2-4)^2$

$$
\begin{aligned}
\text{s.t.} \quad x_1 + \quad x_2 &\geqslant 5 \\
-x_1 \qquad &\geqslant -6 \\
-2x_2 &\geqslant -11 \\
\text{且} \quad x_1, x_2 &\geqslant 0
\end{aligned}
$$

本例与上题的区别仅在于约束条件部分有所不同。考虑到它是线性约束,可行区域又是实心多边形,但对于等值圆来说,这种新的几何位置会产生一个新的结果。正如图 21.1$b$ 所说明的那样,自由极小值解点(4,4)这时在可行集之内,因而就在这一点找到了约束最优解,且有 $\bar{C}=0$。因此,本例中的最优解甚至不在可行区域的边界上,因而最优解一个约束条件也不满足。与线性规划的情况相反,现在再也不可能把选择区域缩小到可行区域的极点集合了。

**例 3** Max. $\pi = 2x_1 + x_2$

$$
\begin{aligned}
\text{s.t.} \quad -x_1^2 + 4x_1 - x_2 &\leqslant 0 \\
2x_1 + 3x_2 &\leqslant 12 \\
\text{且} \quad x_1,\ x_2 &\geqslant 0
\end{aligned}
$$

在本例中,非线性通过第一个约束条件进入模型。把它改写成 $x_2 \geqslant -x_1^2 + 4x_1$,此处右边表达式为 $x_1$ 的二次函数,我们看到这个约束条件要求我们只选择图 21.2 所示的抛物线上或其上方的点。另一方面,第二个约束条件要求所求的点位于斜率为负的直线上或其下方。因此,可行区域总共由不相交的两部分 $F_1$ 和 $F_2$ 所组成。所以,在这种情况下可行区域甚至不是一个凸集!

由线性目标函数,我们得到一族线性等值曲线。就 $F_1$ 而言,在点 P 得出$\pi$ 的最大值,但因 $F_2$ 也是可行区域,点 $P$ 只能称为局部最优值,不是整体最优值。事实上,选择 $F_2$ 中任意一点均比点

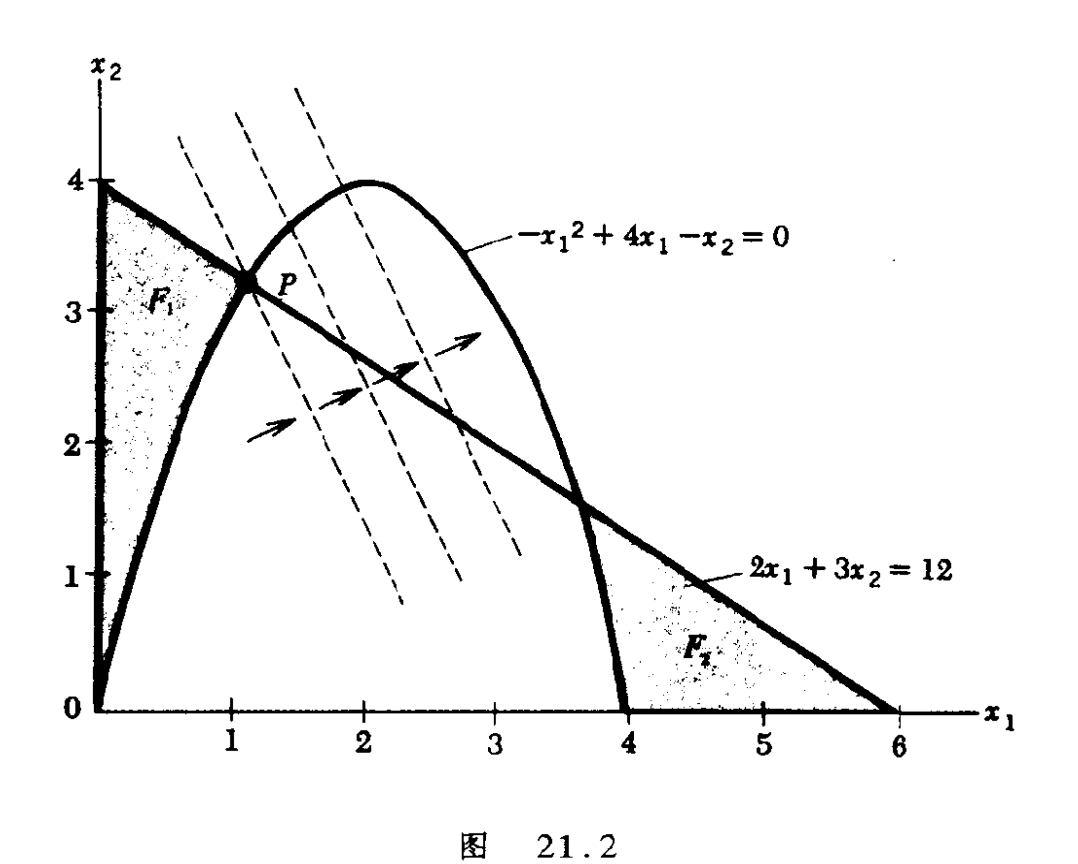



图　21.2

$P$ 都要好。这足以说明:当可行集不是凸集时,就不满足整体性定理(19.3 节)的充分条件。因此,局部最优值不一定是整体最优值。

总之,非线性规划与线性规划的不同之处至少有下列五个方面,其中有些是彼此密切相关的:(1)选择范围扩充到整个可行区域,不只是极点的集合;(2) 所满足的约束条件(和非负限制)的个数不一定恰好等于选择变量的个数;(3) 沿着不变方向移动,不一定能得到目标函数连续的递增(递减)值;(4) 可行区域可能不是凸集;(5) 局部最优值可能不是整体最优值。由于这些差别,适用于线性规划的解法,在非线性的结构中,在很大程度上已不适用了,需要新的解法。但在本章我们的注意力主要不是集中在解的算法上(它往往比较复杂,需要专门研究),而是集中在某些分析结构(必要条件和充分条件)上。这些结果提供最优解的定性特征而

721

不是定量解的本身。

## 练习 21.1

用图解法解下列三个非线性规划问题。在每种情况下，求出具体值 $\bar{x}_1$ 和 $\bar{x}_2$：

1 Min. $C = x_1^2 + x_2^2$

s.t. $x_1 x_2 \geqslant 25$

且 $x_1, x_2 \geqslant 0$

2 Max. $\pi = x_1^2 + (x_2 - 2)^2$

s.t. $5x_1 + 3x_2 \leqslant 15$

且 $x_1, x_2 \geqslant 0$

3 Min. $C = x_1 + x_2$

s.t. $x_1^2 + x_2 \geqslant 9$

$-x_1 x_2 \geqslant -8$

且: $x_1, x_2 \geqslant 0$

722 4 某厂对其第一种产品的需求函数为线性需求函数 $x_1 = a - bP_1$，对第二种产品的线性需求函数为 $x_2 = c - dP_2$。如果这两种产品的平均可变成本分别为 $V_1 = m + x_1$ 和 $V_2 = n + x_2^2$，求其总利润目标函数。

5 文中例 1 和例 2 的目标函数 $C = (x_1 - 4)^2 + (x_2 - 4)^2$，在 $x_1x_2$ 平面中生成一族同心圆。如果我们引进第三个维度 $C$，垂直于 $x_1x_2$ 平面，那么目标函数将产生何类型的曲面？

6 把练习 21.1－1 的非线性规划用下列方法变成等价的古典约束最优化问题：(1) 使用虚拟变量 $s$，(2) 把 $x_1, x_2$ 和 $s$ 分别表示成另外三个变量 $u$，$v$ 和 $w$ 的平方。用古典方法解这个问题，并将其解与前面得到的图解进行比较。

## 21.2 库恩—塔克条件

在古典最优化问题中，对选择变量的符号没有明显的限制，且约束条件中也没有出现不等式。相对或局部极值的一阶条件就是目标函数对所有选择变量及拉格朗日乘数的一阶偏导数等于零。在非线性规划中存在类似的一阶条件，通称为库恩—塔克条件(*the Kuhn - Tucker conditions*)[①]。正如我们将看到的，虽然古典的一阶条件总是必要条件，但库恩—塔克条件不一定是必要条件，除非满足某些附加条件。另一方面，在某些特殊情况下，库恩—塔克条件会变成充分条件，甚至会变成充分必要条件。

因为库恩—塔克条件是非线性规划中唯一最重要的分析结果，所以适当地了解这些条件及其含义是必要的。为了便于说明，我们分两步推出这些条件。

### 非负限制的作用

第一步，我们考察具有非负限制但没有其他约束条件的问题。考察单变量的情况，得到

$$\text{Max.} \quad \pi = f(x_1)$$
$$\text{s.t.} \quad x_1 \geqslant 0$$

其中假设 $f$ 是可微函数。由于 $x_1 \geqslant 0$ 的限制，产生下面三种可能的情况。第一，如果$\pi$的局部极大值出现在图 21.3 中的阴影可行 723

① H.W.库恩和 A.W.塔克:《非线性规划》，收集在 J.内曼(J.Neyman)主编的关于数理统计和概率论的第二次伯克利(Berkiley)专题讨论会论文集中，加利福尼亚大学出版社，1951 年，第 481—492 页。

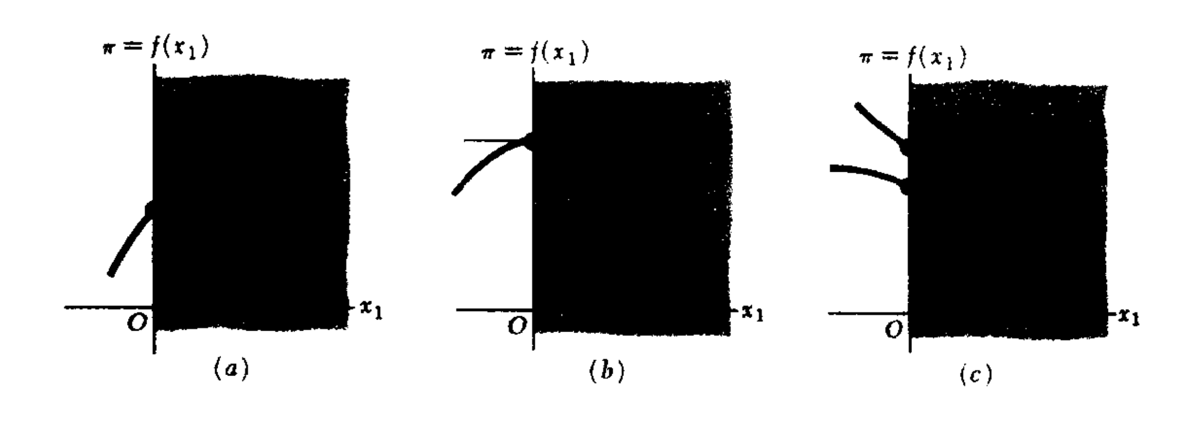



图 21.3

区域内部,例如,出现在图 $a$ 中的 $A$ 点,那么得到一个内解(*interior solution*)。在这种情况下,一阶条件是 $d\pi/dx_1=f'(x_1)=0$,这同古典问题中的一阶条件是一样的。第二,局部极大值也可能出现在纵轴上,如图 $b$ 中点 $B$ 所说明的情况。这里 $x_1=0$。即使在得到边界解的第二种情况下,一阶条件 $f'(x_1)=0$ 仍然成立。作为第三种可能性,局部极大值这时可在图 $c$ 中点 $C$ 或点 $D$ 达到。因为作为规划问题(21.3)局部极大值点的候选点,仅须高于可行区域内邻近的点。考虑到最后一种可能性,像(21.3)这类问题极大值点的特征,不仅由方程 $f'(x_1)=0$,而且由不等式 $f'(x_1)<0$ 来表示。另一方面要注意,方向相反的不等式确实被排除了,因为在向上倾斜的曲线上的点,根本不能得到极大值,即使该点像图 $a$ 中的点 $E$ 那样,位于纵轴上,也是一样的。

上面的讨论的要点是:在问题(21.3)中为使$\pi$在值 $x_1$ 处得到局部极大值,必满足下列三个条件之一:

$$f'(x_1)=0 \quad 和 \quad x_1>0 \quad [点\ A] \tag{21.4}$$

$$f'(x_1)=0 \quad 和 \quad x_1=0 \quad [点\ B] \tag{21.5}$$

$$f'(x_1)<0 \quad 和 \quad x_1=0 \quad [点\ C\ 和\ D] \tag{21.6}$$

实际上,这三个条件可合并一:

$$f'(x_1) \leqslant 0 \qquad x_1 \geqslant 0 \quad \text{和} \quad x_1 f'(x_1) = 0 \qquad (21.7)$$

(21.7)中第一个不等式概括了列在(21.4)到(21.6)中的关于$f'(x_1)$的情况。第二个不等式对 $x_1$ 作类似的归纳。事实上,它仅是非负限制的重述。(21.7)的第三部分是一个方程,它表示了(21.4),(21.5)以及(21.6)共同的特征,即两个量 $x_1$ 与 $f'(x_1)$中 724 至少有一个必取零值,从而使其积为零。(21.7)的三部分构成了选择变量为非负的问题的局部极大值的一阶必要条件。但是深入一步,也可将其视为整体极大值的必要条件。这是因为整体极大值必须也是局部极大值,因此也必满足局部极大值的必要条件。

当含有 $n$ 个选择变量时,问题变为:

$$\begin{aligned} &\text{Max.} \quad \pi = f(x_1, x_2, \cdots, x_n) \qquad (21.8) \\ &\text{s.t.} \quad x_j \geqslant 0 \ (j = 1, 2, \cdots, n) \end{aligned}$$

同样,古典一阶条件 $f_1 = f_2 = \cdots = f_n = 0$ 必需作类似的修改。为此,我们可把(21.7)所依赖的推理逐个应用到每一个选择变量 $x_j$ 上。在图形上,这就相当于把图 21.3 中的横轴依次看成表示每个 $x_j$。这样,容易知道所要求的修正的一阶条件为:

$$f_j \leqslant 0 \quad x_j \geqslant 0 \qquad \text{和} \qquad x_j f_j = 0 \ (j = 1, 2, \cdots, n) \qquad (21.9)$$

其中,$f_j$,是偏导数 $\partial\pi/\partial x_j$。

## 不等式约束的作用

有了这些基本知识,现在进行第二步,力图同时包含不等式约束。为简单起见,首先讨论含三个选择变量($n = 3$)和两个约束($m = 2$)条件的问题:

$$\begin{aligned}
\text{Max.} \quad & \pi = f(x_1, x_2, x_3) \\
\text{s.t.} \quad & g^1(x_1, x_2, x_3) \leqslant r_1 \qquad (21.10) \\
& g^2(x_1, x_2, x_3) \leqslant r_2 \\
\text{且} \quad & x_1, x_2, x_3 \geqslant 0
\end{aligned}$$

利用两个虚拟变量 $s_1$ 和 $s_2$,可把上述问题转化为等价的形式:

$$\begin{aligned}
\text{Max.} \quad & \pi = f(x_1, x_2, x_3) \\
\text{s.t.} \quad & g^1(x_1, x_2, x_3) + s_1 = r_1 \qquad (21.10') \\
& g^2(x_1, x_2, x_3) + s_2 = r_2 \\
\text{且} \quad & x_1, x_2, x_3, s_1, s_2 \geqslant 0
\end{aligned}$$

725 如果没有非负约束,与古典解法一样,作出拉格朗日函数(用 $y$ 而不用 $\lambda$ 表示拉格朗日乘数):

$$\begin{aligned}
Z^* = & f(x_1, x_2, x_3) + y_1[r_1 - g^1(x_1, x_2, x_3) - s_1] \\
& + y_2[r_2 - g^2(x_1, x_2, x_3) - s_2] \qquad (21.11)
\end{aligned}$$

写出一阶条件为:

$$\frac{\partial Z^*}{\partial x_1} = \frac{\partial Z^*}{\partial x_2} = \frac{\partial Z^*}{\partial x_3} = \frac{\partial Z^*}{\partial s_1} = \frac{\partial Z^*}{\partial s_2} = \frac{\partial Z^*}{\partial y_1} = \frac{\partial Z^*}{\partial y_2} = 0$$

但因为变量 $x_j$ 和 $s_i$ 确实必须为非负,所以,对于这些变量的一阶条件应按照(21.9)加以修正。结果,我们得到如下一组条件:

$$\begin{aligned}
& \frac{\partial Z^*}{\partial x_j} \leqslant 0 \qquad x_j \geqslant 0 \qquad \text{和} \qquad x_j \frac{\partial Z^*}{\partial x_j} = 0 \\
& \frac{\partial Z^*}{\partial s_i} \leqslant 0 \qquad s_i \geqslant 0 \qquad \text{和} \qquad s_i \frac{\partial Z^*}{\partial s_i} = 0 \qquad (21.12) \\
& \frac{\partial Z^*}{\partial y_i} = 0 \qquad \text{和} \qquad \begin{pmatrix} i = 1,2 \\ j = 1,2,3 \end{pmatrix}
\end{aligned}$$

注意,导致 $\partial Z^* / \partial y_i$ 确实仍被确定为零。(为什么?)

(21.12)的各行与不同类型的变量相联系。但我们可把后两行合并,并在合并过程中由一阶条件消去虚拟变量 $s_i$。由于$\partial Z^* = \partial s_i = -y_i$,由第二行可知,必有 $-y_i \leqslant 0$, $s_i \geqslant 0$ 和 $-s_i y_i = 0$ 或等价地

$$s_i \geqslant 0 \qquad y_i \geqslant 0 \quad 和 \quad y_i s_i = 0 \tag{21.13}$$

但第三行[(21.10′)中约束条件的复述]意味着

$$s_i = r_i - g^i(x_1, x_2, x_3)。$$

因此,把它代入(21.13),可把(21.12)的第二和第三行合并为:

$$r_i - g^i(x_1, x_2, x_3) \geqslant 0 \qquad y_i \geqslant 0$$

$$和 \qquad y_i[r_i - g^i(x_1, x_2, x_3)] = 0$$

这使我们不用虚拟变量把一阶条件(21.12)用等价的形式表示出来。用符号 $g_j^i$ 表示$\partial g^i / \partial x_j$,现在可写成

$$\frac{\partial Z^*}{\partial x_j} = f_j - (y_1 g_j^1 + y_2 g_j^2) \leqslant 0 \qquad x_j \geqslant 0 \quad 和 \quad x_j \frac{\partial Z^*}{\partial x_j} = 0$$

$$r_i - g^i(x_1, x_2, x_3) \geqslant 0 \ y_i \geqslant 0 \quad 和 \quad y_i[r_i - g^i(x_1, x_3, x_3)] = 0 \tag{21.14}$$

这是问题(21.10)的库恩—塔克条件,或更准确地说,它是库恩—塔克条件的一种形式,这种形式是用(21.11)中的拉格朗日函数 $Z^*$ 表示的。

尽管也有可能使用不同的拉格朗日函数更直接提到同样的一组条件,但现在我们还是知道了这个结果。假设规划问题是(21. 726 10),我们忽略非负限制和约束中的不等式,写出拉格朗日函数 $Z$ 的纯古典形式:

$$Z = f(x_1, x_2, x_3) + y_1[r_1 - g^1(x_1, x_2, x_3)] + y_2[r_2 - g^2(x_1, x_2, x_3)] \tag{21.15}$$

则我们(1)规定偏导数$\partial Z/\partial x_j \leqslant 0$,但$\partial Z/\partial y_i \geqslant 0$;(2) 对 $x_j$ 和 $y_i$ 加上非负限制;(3) 要求每个变量和 $Z$ 对该变量的偏导数之间有互补松弛关系,即要求它们的积为零。因这两步的结果,即

$$\frac{\partial Z}{\partial x_j} = f_j - (y_1 g_j^1 + y_2 g_j^2) \leqslant 0 \quad x_j \geqslant 0 \quad \text{和} \quad x_j \frac{\partial Z}{\partial x_j} = 0$$

$$\frac{\partial Z}{\partial y_i} = r_i - g^i(x_1, x_2, x_3) \geqslant 0 \quad y_i \geqslant 0 \quad \text{和} \quad y_i \frac{\partial Z}{\partial y_i} = 0$$

(21.16)

与(21.14)是一致的,所以库恩—塔克条件也可用拉格朗日函数 $Z$(与 $Z^*$ 对比)来表示。注意,从 $Z^*$ 转到 $Z$ 时,我们不仅更直接地得到库恩—塔克条件,而且使表达式 $r_i - g^i(x_1, x_2, x_3)$[在(21.14)中没有命名的部分]等于偏导数。因此,在下面的讨论中我们只使用建立在拉格朗日函数 $Z$ 基础上的库恩—塔克条件的(21.16)形式。

### 库恩—塔克条件的解释

库恩—塔克条件(21.16)中的一部分仅是给定问题中某些方面的重述。例如,条件 $x_j \geqslant 0$ 仅是非负限制的复述,条件$\partial Z/\partial y_i \geqslant 0$ 仅是约束条件的重复。但将其包含(21.16)中,对于更明确地揭示两种变量 $x_j$(选择变量)与 $y_i$(拉格朗日乘数)之间值得关注的对称性非常有利。每一类型的各个变量都有最优解满足对应的边际条件($\partial Z/\partial x_j \leqslant 0$ 或$\partial Z/\partial y_i \geqslant 0$)。每个变量也必为非负。最后,每个变量都具有与拉格朗日函数 $Z$ 的某个特定偏导数具有互补松弛性的特征。这表明,对每个 $x_j$,在最优解的情况下必有:或者边际条件像在古典结构中那样作为等式成立,或者所研究的选择

变量取零值，或者两种情况都存在。同样，对每个 $y_i$，在最优解的情况下必得到：或者边际条件作为等式成立（其含义是满足第 $i$ 个约束条件的等式约束），或者拉格朗日乘数等于零，或者两种情况都发生。 727

当我们考察(21.16)中的 $\partial Z/\partial x_j$ 和 $\partial Z/\partial y_i$ 的展式时，甚至还能够给出更明确的解释。假设此问题是我们熟悉的生产问题，那么得到

$f_j \equiv$ 第 $j$ 种产品的边际总利润

$y_i \equiv$ 第 $i$ 种资源的影子价格

$g_j^i \equiv$ 生产第 $j$ 种产品的边际单位所使用的第 $i$ 种资源的数量。

$y_i g_j^i \equiv$ 生产第 $j$ 种单位产品需要的第 $i$ 种资源的边际投入成本

$\sum_i y_i g_j^i \equiv$ 第 $j$ 种产品的总边际投入成本。

这样，边际条件

$$\frac{\partial Z}{\partial x_j} = f_j - \sum_i y_i g_j^i \leqslant 0$$

要求第 $j$ 种产品的总边际总利润不大于其总边际投入成本，即投入不足是不可以的。因而互补松弛条件表明：如果最优解要求第 $j$ 种产品积极生产($\bar{x}_j > 0$)，其边际总利润恰等于总边际估算成本($\partial Z/\partial \bar{x}_j = 0$)，这同古典最优化问题的情况是一致的。另一方面，如果最优的边际总利润小于总投入成本($\partial Z/\partial \bar{x}_j < 0$)，即承担**超额**投入，则这种产品不可以生产($\bar{x}_j = 0$)。[①] 后一种情况在古典结构中是不会出现的，因为如果边际总利润小于边际投入成本，那末产

① 注意，已知方程 $ab=0$，此处 $a$ 和 $b$ 是实数，我们可以正确地推断：$a \neq 0$ 必有 $b=0$。但 $a=0$ 则 $b \neq 0$ 这个命题不真，因 $b=0$ 时也可有 $a=0$。

品在这种结构中会一直减少，直到边际利润与边际成本相等。这里，导致$\partial Z/\partial x_j<0$作为最优化条件的，是在这种结构中的明确的非负性设定。因为那时我们在降低产出方面所能做到的就是把生产减少到零水平。如果在零的产出水平上仍然得到$\partial Z/\partial \bar{x}_j<0$，那也就无能为力了。

至于其余的条件，都是与变量 $y_i$ 有关，其含义更容易看出。首先，边际条件$\partial Z/\partial y_i\geqslant 0$仅要求厂商将该厂将每种资源的生产能力保持在这个限度以内。互补松弛条件规定，如果第 $i$ 种资源在最优解的情况下，没有充分使用($\partial Z/\partial \bar{y}_i>0$)，那么这种资源的影子价格(不允许为负)必然等于零($\bar{y}_i=0$)。另一方面，如果在最优解的情况下资源的影子价格为正($\bar{y}_i>0$)，那么这种资源必定是一种充分利用的资源($\partial Z/\partial \bar{y}_i=0$)。当然，这不过是上一章对偶定理II的含义而已。

当然，也可以把拉格朗日乘数值 $\bar{y}_i$ 看作度量目标函数最优值对第 $i$ 种约束条件的轻微松弛是如何作出反应的指标(见12.2节)。从这方面看，互补松弛意味着如果第 $i$ 种约束条件不受最优化约束($\partial Z/\partial \bar{y}_i>0$)，那么放松这个特殊约束条件并不影响总利润的最优值($\bar{y}_i=0$)就像放松腰上没有捆紧的皮带，不产生任何更舒服的感觉一样。另一方面，如果稍微松弛第 $i$ 个约束条件(使第 $i$ 种资源的资金增加)的确能增加总利润($\bar{y}_i>0$)，那么事实上这种资源的约束条件在最优解的情况下是有约束力的($\partial Z/\partial \bar{y}_i=0$)。

### $n$ 个变量 $m$ 个约束的情况

上面的讨论可以直接推广到(21.1)或(21.1′)中所给出的最

优化问题中。因为现在有 $n$ 个选择变量和 $m$ 个约束条件，拉格朗日函数以下列更为一般的形式出现：

$$Z = f(x_1, x_2, \cdots, x_n) + \sum_{i=1}^{m} y_i [r_i - g^i(x_1, x_2, \cdots, x_n)] \tag{21.17}$$

而库恩—塔克条件则是

$$\begin{aligned} &\frac{\partial Z}{\partial x_j} \leqslant 0 \qquad x_j \geqslant 0 \quad \text{和} \quad x_j \frac{\partial Z}{\partial x_j} = 0 \\ &\frac{\partial Z}{\partial y_i} \geqslant 0 \qquad y_i \geqslant 0 \quad \text{和} \quad y_i \frac{\partial Z}{\partial y_i} = 0 \qquad [\text{极大化}] \end{aligned} \tag{21.18}$$

$$\begin{pmatrix} i = 1, 2, \ldots, m \\ j = 1, 2, \cdots, n \end{pmatrix}$$

这里，为了避免出现混乱，我们没有写出偏导数 $\partial Z / \partial x_j$ 和 $\partial Z / \partial y_i$ 的展开式。但为了更详细地考察类似于(21.16)中所给出的库恩—塔克条件，还是要求读者把其展开式写出来。注意，除了问题的维数变化以外，库恩—塔克条件完全同以前讨论过的三个变量，两个约束条件的情况一样。当然，对这些条件的解释也仍然一样。

如果规划问题是像(21.2)和(21.2′)中那样的极小化问题，情况会怎样呢？一种处理的方法是把它化为极大化问题，然后应用(21.18)。使 $C$ 极小等价于使 $-C$ 极大，所以这样的转换总是可行的。当然，我们要用 $(-1)$ 通乘每个约束条件以变换约束不等式。但我们也可以不用变换的方法，而再次使用(21.17)中所定义的拉格朗日函数，直接应用下列库恩—塔克条件的极小化形式：

729

$$\frac{\partial Z}{\partial x_j} \geqslant 0 \qquad x_j \geqslant 0 \quad \text{和} \quad x_j \frac{\partial Z}{\partial x_j} = 0 \tag{21.19}$$

$$\frac{\partial Z}{\partial y_i} \leqslant 0 \qquad y_i \geqslant 0 \quad 和 \quad y_i \frac{\partial Z}{\partial y_i} = 0 \quad [极小化]$$

$$\begin{pmatrix} i = 1, 2, \ldots, m \\ j = 1, 2, \cdots, n \end{pmatrix}$$

读者应将其与(21.18)比较。

按水平方向(行的方向)读出(21.18)和(21.19)时,我们看到极大化和极小化问题的库恩—塔克条件,分别由与选择变量 $x_j$(第 1 行)有关的一组条件以及另外一组与拉格朗日乘数 $y_i$(第 2 行)相关的条件所构成。另一方面,按纵向(列的方向)读出时,我们应注意到,对每一 $x_j$ 和 $y_i$,存在一个边际条件(第一列)、非负约束(第二列)和互补松弛条件(第 3 列)。在任意已知的规划问题中,与选择变量有关的一组边际条件同拉格朗日乘数的一组边际条件的不同之处在于其不等式所取的方向。同样,极大化问题的一组边际条件与极小化问题的边际条件的差别也在于其不等式所取的方向。

如果满足下一节所说明的条件,那么库恩—塔克极大条件(21.18)和极小条件(21.19)分别是局部极大化和局部极小化的必要条件。但因整体极大(极小)也必是局部极大(极小),所以如果满足同一条件的话,库恩—塔克条件也可作为整体极大(小)的必要条件。

## 一个例子

我们来检验如图 21.1$a$ 所说明的 21.1 节中例 1 的解是否满足库恩—塔克条件。该问题的拉格朗日函数是

$$Z = (x_1 - 4)^2 + (x_2 - 4)^2 + y_1(6 - 2x_1 - 3x_2)$$

$$+ y_2(-12 + 3x_1 + 2x_2)$$

因为这个问题是一个极小化问题,适当的条件是(21.19),它包含下列四个边际条件: 730

$$\frac{\partial Z}{\partial x_1} = 2(x_1 - 4) - 2y_1 + 3y_2 \geqslant 0$$

$$\frac{\partial Z}{\partial x_2} = 2(x_2 - 4) - 3y_1 + 2y_2 \geqslant 0$$

$$\frac{\partial Z}{\partial y_1} = 6 - 2x_1 - 3x_2 \leqslant 0$$

$$\frac{\partial Z}{\partial y_2} = -12 + 3x_1 + 2x_2 \leqslant 0$$

加上非负条件以及互补松弛条件。问题是:我们能求出非负值 $\bar{y}_1$ 和 $\bar{y}_2$,它们与最优值 $\bar{x}_1 = 2\frac{2}{13} = \frac{28}{13}$和 $\bar{x}_2 = 2\frac{10}{13} = \frac{36}{13}$一起,满足所有这些条件吗?

给定 $\bar{x}_1$ 和 $\bar{x}_2$ 都为非零,互补松弛表明 $\partial Z/\partial x_1 = 0$ 和 $\partial Z/\partial x_2 = 0$。于是,将 $\bar{x}_1$ 和 $\bar{x}_2$ 值代入前两个边际条件,它们变为两个方程:

$$-2y_1 + 3y_2 = \frac{48}{13}$$

$$-3y_1 + 2y_2 = \frac{32}{13}$$

其解为 $\bar{y}_1 = 0$ 和 $\bar{y}_2 = \frac{16}{13} = 1\frac{3}{13}$, 均为非负,这正是所要求的。因这些值边同 $\bar{x}_1$ 和 $\bar{x}_2$ 一起,意味着$\partial Z/\partial \bar{x}_1 = 0$,$\partial Z/\partial \bar{x}_2 = 0$,$\frac{\partial Z}{\partial \bar{y}_1} < 0$ 和$\partial Z/\partial \bar{y}_2 = 0$,这些值满足边际条件和互补松弛条件,所以满足库恩—塔克极小化条件。

## 练习 21.2

1　对于极小化的情况，画出类似于图 21.3 的一组图形，并推导一组对应于(21.4)到(21.6)的局部极小化的必要条件。然后把这些条件简化为类似于(21.7)的单个陈述。

2　(*a*)证明：在(21.18)中，一组 $m$ 个独立的条件不必写成

$$y_i \frac{\partial Z}{\partial y_i} = 0 \qquad (i = 1,2,\cdots,m)$$

的形式，写出形式为

$$\sum_{i=1}^{m} y_i \frac{\partial Z}{\partial y_i} = 0$$

的单个方程便足矣。

(*b*) 对下列一组条件

$$x_j \frac{\partial Z}{\partial x_j} = 0 \qquad (j = 1,2,\cdots,n)$$

我们能像前面一样改写吗？

731　3　利用前题的推理，(21.19)中哪一(几)组条件可简化为单个的方程？

4　已知极小化问题(21.2)，利用拉格朗日函数(21.17)，求出导数 $\partial Z/\partial x_j$ 和 $\partial Z/\partial y_i$，并写库恩—塔克极小化条件(21.19)的展开式。

5　把极小化问题(21.2)变为极大化问题，列出拉格朗日函数的表达式，对 $x_j$ 和 $y_i$ 取导数，并应用库恩—塔克极大条件(21.18)。这些结果与在前一问题中所得到的结果一致吗？

6　将库恩—塔克条件应用到 21.1 节中的例 2，并作如下检验：

(*a*)写出拉格朗日函数和库恩—塔克条件。

(*b*)由图 21.1*b* 中所给出的解，求出 $\partial Z/\partial y_i$ ($i = 1,2,3$) 的最优值。那么关于 $\bar{y}_i$，我们能得出什么结论？

(*c*)求出 $\partial Z/\partial x_1$ 和 $\partial Z/\partial x_2$ 的最优值。

(*d*)所有库恩—塔克条件都满足吗？

# 21.3 约束规范

上一节已指出，只有满足特定条件时，库恩—塔克条件才是必要条件。这条件叫做**约束规范**（*constraint qualification*）。约束规范是对非线性规划中的约束函数施加的某些限制，目的是为了排除可行集边界上的某些不规则性，这些不规则性可能会违背能够产生最优解的库恩—塔克条件。

## 边界点的不规则性

我们先用一些具体例子来说明这种不规则性的性质

**例 1**　Max.　$\pi = x_1$

s.t.　$x_2 - (1 - x_1)^3 \leqslant 0$

和　$x_1,\ x_2 \geqslant 0$

如图 21.4 所示，可行区域是第一象限中位于曲线 $x_2 = (1 - x_1)^3$ 732 上及其下方的点的集合。

因为目标函数告诉我们，要使 $x_1$ 最大，最优解是点(1,0)。但这个解不满足库恩—塔克极大化条件。为了检验这一点，我们首先写出拉格朗日函数

$$Z = x_1 + y_1[-x_2 + (1 - x_1)^3]$$

作为第一边际条件，我们应得到

$$\frac{\partial Z}{\partial x_1} = 1 - 3y_1(1 - x_1)^2 \leqslant 0$$

事实上，因为 $\bar{x}_1 = 1$ 为正，当在点(1,0)计值时，互补松弛要求该导数等于零。但我们得到的实际值恰巧为$\partial Z / \partial \bar{x}_1 = 1$，因此，违背

了上面的边际条件。

这种异常产生的原因在于本例的最优解(1,0)出现在向外指的歧点(*cusp*)。它构成了使库恩—塔克条件在边界的最优解失效的一种不规则性。当曲线突然反向,使在该点一边的斜率等于该点另一边的斜率时,歧点就变成尖点(*sharp point*)。这里,可行区域的边界首先顺着约束曲线行走,当到达点(1,0)时,突然向西转,此后,再沿着水平方向前进。因为在点(1,0)处弯曲的边界和水平的边界的斜率都是零,所以该点为歧点。

歧点是使库恩—塔克条件失效最经常引用的原因,但事实上歧点的出现既不是库恩—塔克条件在最优解失效的必要条件,也不是充分条件。下面两个例子就可证实这一点。

**例 2** 对于前一例子的问题,我们加上新的约束条件

$$2x_1 + x_2 \leqslant 2$$

其边界为 $x_2 = 2 - 2x_1$,在图 21.4 中绘出其图形是一条斜率为 $-2$ 且通过原点的直线。显然,可行区域仍然同以前一样,最优解也出现在歧点。但如果我们写出新的拉格朗日函数

$$Z = x_1 + y_1[-x_2 + (1 - x_1)^3] + y_2[2 - 2x_1 - x_2]$$

733 和边际条件

$$\frac{\partial Z}{\partial x_1} = 1 - 3y_1(1 - x_1)^2 - 2y_2 \leqslant 0$$

$$\frac{\partial Z}{\partial x_2} = -y_1 - y_2 \leqslant 0$$

$$\frac{\partial Z}{\partial y_1} = -x_2 + (1 - x_1)^3 \geqslant 0$$

$$\frac{\partial Z}{\partial y_2} = 2 - 2x_1 - x_2 \geqslant 0$$

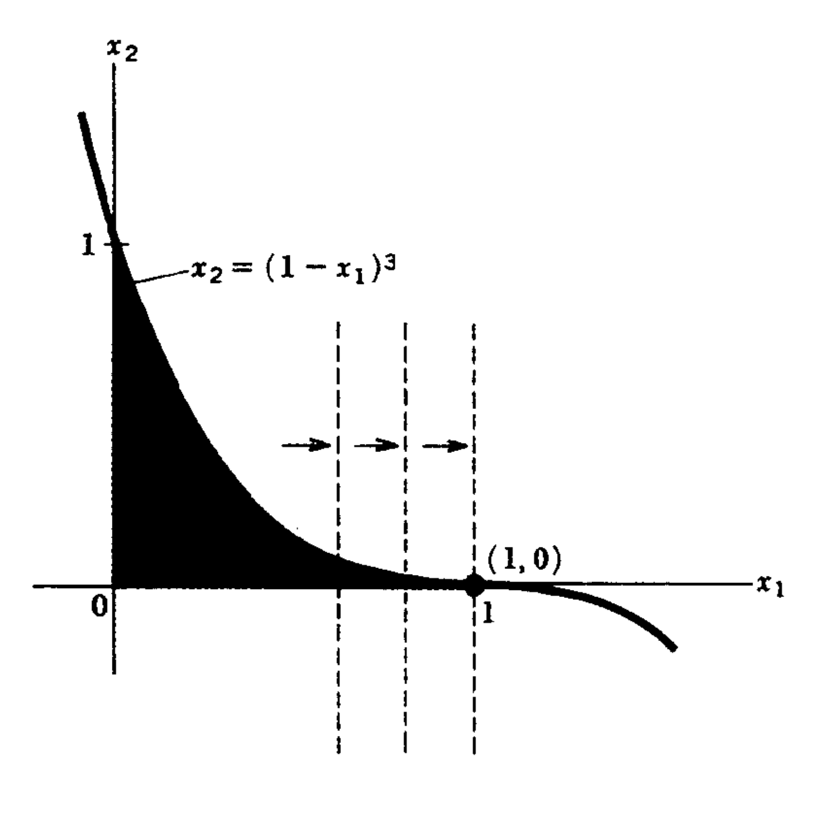



图 21.4

这表明 $\bar{x}_1=1,\bar{x}_2=0,\bar{y}_1=1$ 及 $\bar{y}_2=\frac{1}{2}$ 的确满足上面四个不等式,而且也满足非负限制和互补松弛条件。事实上,$y_1$ 可被赋予任意非负值(不只等于 1),所有条件仍然满足。这就说明了拉格朗日乘数的最优值不一定是唯一的。但重要的是,本例说明尽管有歧点,但库恩—塔克条件仍然成立。

**例 3** 问题

$$
\begin{aligned}
\text{Max.}\quad & \pi = x_2 - x_1^2 \\
\text{s.t.}\quad & -(10-x_1^2-x_2)^3 \leqslant 0 \\
& -x_1 \leqslant -2 \\
\text{和}\quad & x_1, x_2 \geqslant 0
\end{aligned}
$$

的可行区域如图 21.5 所示,任何地方都不含歧点。但在最优解(2,6),库恩—塔克条件仍然不成立。拉格朗日函数为

$$Z = x_2 - x_1^2 + y_1(10 - x_1^2 - x_2)^3 + y_2(-2 + x_1)$$

第二个边际条件要求

$$\frac{\partial Z}{\partial x_2} = 1 - 3y_1(10 - x_1^2 - x_2)^2 \leqslant 0$$

因 $\bar{x}_2$ 为正,这个导数在点(2,6)计值时,确实等于零。但无论 $y_1$ 被赋予何值,实际上都得到 $\partial Z/\partial \bar{x}_2 = 1$。因此,库恩—塔克条件甚至在没有歧点的情况下,亦即当可行区域是图21.5中的凸集时,也可能不成立。对于库恩—塔克条件的失效,歧点的存在既不是必要条件,也不是充分条件,其基本原因在于上面所提到的不规则性与可行区域本身的形状无关,而与约束函数本身的形式有关。

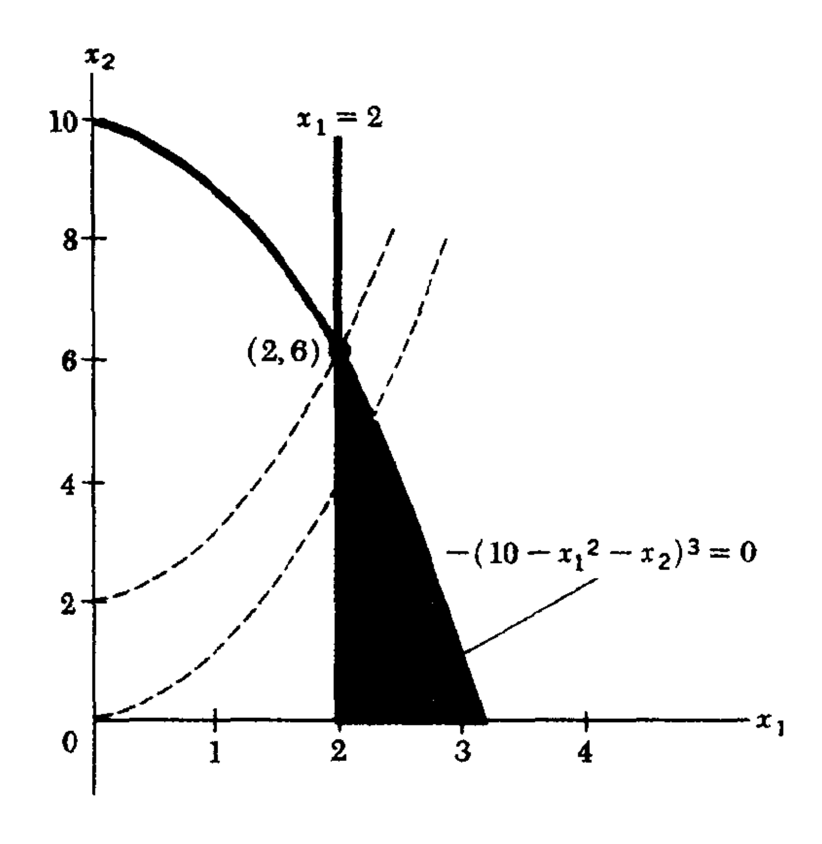



图 21.5

如果满足某一约束规范,则边界的不规则性(有歧点或没有歧点的不规则性)就不可能出现。

为了解释这一点,令 $\bar{x}\equiv(\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_n)$是可行区域边界上的一个(解)点,并令 $dx\equiv(dx_1,dx_2,\cdots,dx_n)$表示由所提到的边界点移动的特定方向。将向量 $dx$ 解释为移动的方向,这与我们前面把向量解释为方向线段(箭号)是一致的,但这里的起点是点 $\bar{x}$,而不是原点,因而向量 $dx$ 不具有向量径的性质。现在我们对向量 $dx$ 附加两个要求。第一,如果第 $j$ 个选择变量在点 $\bar{x}$ 处取零值,那么只允许在 $x_j$ 轴上有非负变化,即

$$\text{如果} \quad \bar{x}_j=0 \qquad \text{那么} \quad dx_j\geqslant 0 \qquad (21.20)$$

第二,如果在点 $\bar{x}$ 处恰好满足第 $i$ 个约束条件的等式约束,那么将只允许 $dx_1,\cdots,dx_n$ 的取值使约束函数值 $g^i(\bar{x})$不增加(对极大化问题),或不减少(对极小化问题),即

$$dg^i(\bar{x}) = g_1^i dx_1 + g_2^i dx_2 + \cdots + g_n^i dx_n \begin{cases} \leqslant 0(\text{极大化}) \\ \geqslant 0(\text{极小化}) \end{cases} \qquad (21.21)$$

$$\text{如果} \quad g^i(\bar{x}) = r_i$$

其中所有偏导数 $g_j^i$ 都在 $\bar{x}$ 处计值。如果向量 $dx$ 满足(21.20)和(21.21),则我们称其为**测试向量**(*test vector*)。最后,如果存在满足下列条件的可微弧:(1)从点 $\bar{x}$ 出发;(2)整个包含在可行区域内;(3)与已知测试向量相切,则我们把这样的弧段称为该测试向量的**规范弧**。有了这些预备知识后,约束规范可简单地表述如下: 735

如果对可行区域边界上的任意点 $\bar{x}$，对每一测试向量 $dx$，存在一规范弧，那么，就满足约束规范。

**例4** 证明在图 21.4 中的例 1 的最优点(1,0)不满足库恩—塔克条件，也不满足约束规范。在该点 $\bar{x}_2=0$，于是测试向量满足

$$dx_2 \geqslant 0 \qquad [\text{由}(21.20)]$$

进而，因在点(1,0)，这个(唯一)约束 $g^1 = x_2-(1-x_1)^3 \leqslant 0$ 恰好得到满足，所以必令

$$g_1^1 dx_1 + g_2^1 dx_2 = 3(1-\bar{x}_1)^2 dx_1 + dx_2 = dx_2 \leqslant 0$$

$$[\text{由}(21.21)]$$

这两个要求意味着我们必须令 $dx_2=0$。相反，我们可任选 $dx_1$。因此，向量$(dx_1, dx_2)=(2,0)$，像$(dx_1, dx_2)=(-1,0)$一样，是一个可接受的测试向量。测试向量$(dx_1, dx_2)=(-1,0)$在图21.4中表现为一个由点(1,0)出发并指向正西方向的箭头(没有画出)，显然可以对其绘出一个符合要求的规范弧。(可行区域的弯曲边界可以作为一个规范弧。)另一方面，测试向量$(dx_1, dx_2)=(2,0)$可以绘成一个始于(1,0)并指向正东方向的箭头(没有绘出)。因为没法绘出一个平滑弧与此向量相切并完全位于可行区域内，所以不存在合格的规范弧。因此，最优解点(1,0)违背了约束规范。

**例5** 关于上面的例 2，在把附加约束条件 $2x_1+x_2 \leqslant 2$ 加到图 21.4 中后，我们说明点(1,0)将满足约束规范，从而再证明库恩—塔克条件。

正如在例 4 中那样，我们必要求 $dx_2 \geqslant 0$(因 $\bar{x}_2=0$)和 $dx_2 \leqslant 0$(因满足第一个约束条件的等式约束)；于是，$dx_2=0$。但也满足第二个约束条件的等式约束，因而要求

$$g_1^2 dx_1 + g_2^2 dx_2 = 2dx_1 + dx_2 = 2dx_1 \leqslant 0 \qquad [\text{由}(21.21)]$$

具有非负的 $dx_1$ 和为零的 $dx_2$，唯一可接受的容许测试向量(除零向量本身以外)是从图 21.4 中的点(1,0)出发，指向正东的向量。所有这些向量都位于可行区域的横轴上，对每个测试向量在这里一定能找到规范弧。因此，这一次的确满足约束规范。

## 线性约束条件

前面在例 3 中说明过，可行集的凸性并不能保证库恩—塔克条件能够作为必要条件。但如果可行区域是仅由线性约束形成的凸集，那么约束规范总是满足，且库恩—塔克条件在最优解处总成立。如果是这种情况，讨论具有线性约束条件的非线性规划，或者作为特例，讨论原本就是线性规划的规划问题时，我们就不用担心边界的不规则性了。 736

**例 6** 我们来说明在两个变量和两个线性约束条件的结构中线性约束的结果。对于一个极大化问题，线性约束可以写成

$$a_{11}x_1 + a_{12}x_2 \leqslant r_1$$

$$a_{21}x_1 + a_{22}x_2 \leqslant r_2$$

其中，所有参数取正值。如图 21.6 所示，第一个约束边界的斜率为 $-a_{11}/a_{12}<0$，第二个为 $-a_{21}/a_{22}<0$。阴影可行区域的边界点可分为五类：(1)原点，它是两坐标轴的交点；(2)位于一坐标轴线段上的点，例如点 $J$ 和点 $S$；(3)一轴与一约束边界的交点，即点 $K$ 和点 $R$；(4)位于单个约束边界上的点，如点 $L$ 和点 $N$；(5)两个约束边界的交点 $M$。对于上述各类点的满足约束规范的情况，我们依次进行简单的检验。

1. 在原点，一个约束条件也不满足，因而，我们可以忽略(21.21)。

   但因 $x_1=x_2=0$，由(21.20)，我们必须选择 $dx_1 \geqslant 0$ 和 $dx_2 \geqslant 0$

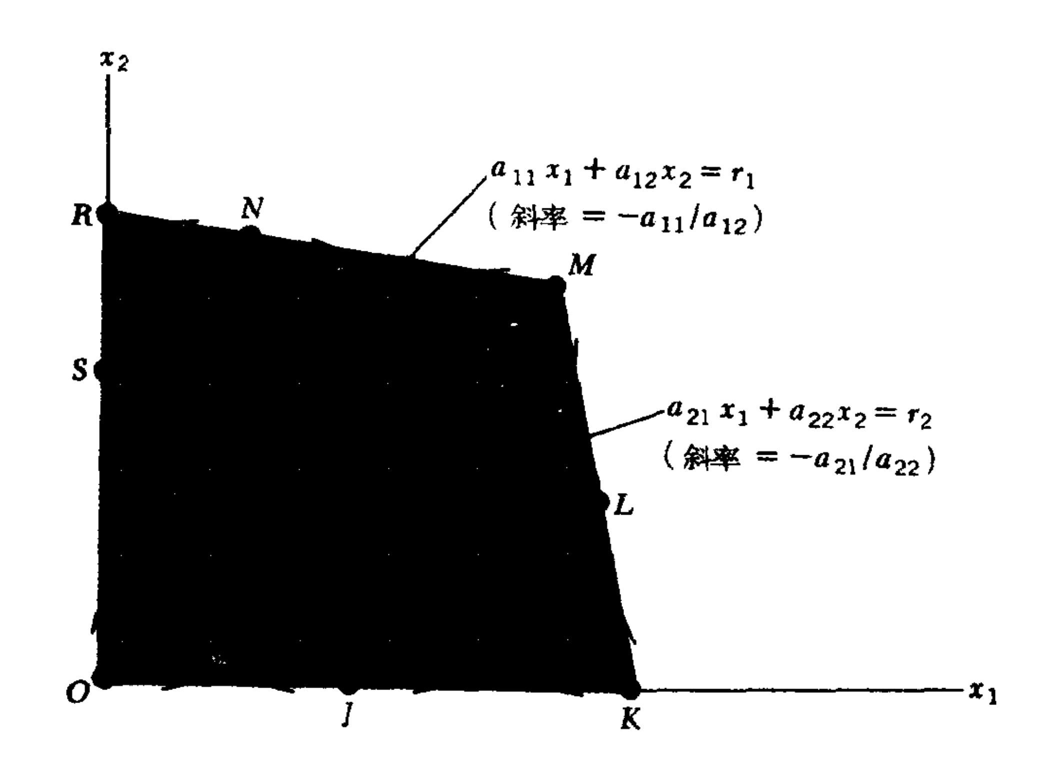



图 21.6

的测试向量。因此,所有从原点出发的测试向量如图 21.6 所描述的,必指向正东、正北或东北方向。这些向量恰好都落在可行集以内,显然,对每个测试向量都可找出规范弧。

737

2. 像在点 $J$ 这样的点,我们又可以忽略(21.21)。$x_2=0$ 表示我们必须选择 $dx_2 \geqslant 0$,但 $dx_1$ 可任意选取。因此,除指向正南的向量($dx_2<0$)外,所有向量都是可接受的测试向量,都落在可行区域内,且对每个测试向量都存在规范弧。对点 $S$ 的分析也是类似的。

3. 在点 $K$ 和点 $R$ 处,必须同时考虑(21.20)和(21.21)。具体地说,在 $K$ 处因 $x_2=0$,我们必选 $dx_2 \geqslant 0$,因而我们必须排除向南的所有箭头。当满足第二个约束条件的等式约束时,点 $K$ 的测试向量必满足

$$g_1^2 dx_1 + g_2^2 dx_2 = a_{21} dx_1 + a_{22} dx_2 \leqslant 0 \quad (21.22)$$

因为在点 $K$ 处也有 $a_{21}x_1 + a_{22}x_2 = r_2$(第二个约束边界),我们可把这个等式加到(21.22),把测试向量的限制修改成如下形式: 738

$$a_{21}(x_1 + dx_1) + a_{22}(x_2 + dx_2) \leqslant r_2 \qquad (21.22')$$

如果把$(x_j + dx_j)$解释为在测试向量箭头处得到的 $x_j$ 的新值,那末可把(21.22′)解释为表示所有测试向量,其箭头位于第二个约束边界上或其下方。因此,所有这些向量也必落在可行区域以内,因而对每个测试向量可求出其规范弧。点 $R$ 的分析也是类似的。

4. 像在 $N$ 和 $L$ 这样的点上,变量都不为零,(21.20)都可以忽略。但对点 $N$,(21.21)表明

$$g_1^1 dx_1 + g_2^1 dx_2 = a_{11}dx_1 + a_{12}dx_2 \leqslant 0 \qquad (21.23)$$

因为点 $N$ 满足 $a_{11}x_1 + a_{12}x_2 = r_1$(第 1 个约束边界),可把这个等式加到(21.23),写成

$$a_{11}(x_1 + dx_1) + a_{12}(x_2 + dx_2) \leqslant r_1 \qquad (21.23')$$

这就要求测试向量的箭头位于图 21.6 中第一个约束边界上或其下方。这样,我们实际上得到在其他几种情况见过的同样结果。点 $L$ 的分析也是类似的。

5. 在点 $M$,我们又可以忽略(21.20),但这一次(21.21)要求所有测试向量同时满足(21.22)和(21.23)。因我们可把后者修改为(21.22′)和(21.23′)的形式。所有测试向量的箭头现在必位于第一和第二个约束边界上或其下方。于是,这结果又同前面的情况完全一样。

在本例中非常碰巧,对所考虑的每种边界点,所有的测试向量都位于可行区域之内。虽然这种位置上的特征使规范弧

容易求出，但决不是规范弧存在的先决条件。特别是，在含有非线性约束边界的问题中，约束边界本身也可用作位于可行区域之内的某个测试向量的规范弧。在练习 21.3 的问题中，可以找到一个这样的例子。

## 练习 21.3

---

1 检验例 3 中的点$(\bar{x}_1,\bar{x}_2)=(2,6)$是否满足约束规范。

2 Max. $\pi = x_1$

s.t. $x_1^2 + x_2^2 \leqslant 1$

且 $x_1, x_2 \geqslant 0$

用图解法解此题。并检验最优解点是否满足

$(a)$约束规范；$(b)$库恩—塔克极大化条件。

3 Min. $C = x_1$

s.t. $x_1^2 - x_2 \geqslant 0$

且 $x_1, x_2 \geqslant 0$

用图解法解此规划。最优解在歧点出现吗？检验最优解是否满足：$(a)$约束规范，$(b)$库恩—塔克极小化条件。

4 Min. $C = 2x_1 + x_2$

s.t. $x_1^2 - 4x_1 + x_2 \geqslant 0$

$-2x_1 - 3x_2 \geqslant -12$

且 $x_1, x_2 \geqslant 0$

用图解法求整体极小值，并检验最优解是否满足：$(a)$约束规范，$(b)$库恩—塔克条件。（提示：可行区域与图 21.2 所描述的一样。）

5 Min. $C = x_1$

s.t. $-x_2 - (1 - x_1)^3 \geqslant 0$

且 $x_1, x_2 \geqslant 0$

证明：$(a)$最优解$(\bar{x}_1,\bar{x}_2)=(1,0)$的确不满足库恩—塔克条件；$(b)$当引进

新乘数 $y_0 \geqslant 0$,把拉格朗日函数修改成如下形式

$$Z_0 = y_0 f(x_1, x_2, \cdots, x_n) + \sum_{i=1}^{m} y_i [r_i - g^i(x_1, x_2, \cdots, x_n)]$$

则在点(1,0)处满足库恩—塔克条件。(注意:关于乘数的库恩—塔克条件只能推广到 $y_1, \cdots, y_m$,不能到 $y_0$。)

---

## 21.4 库恩—塔克充分性定理:凹规划

迄今为止,我们的讨论仅与非线性规划中极大值或极小值的必要条件有关。必要条件作为一种筛选机制,对排除不合格的最优解的候选点是有用的。例如,在非线性规划中满足库恩—塔克条件的可行区域的任意内点,不可能是最优解。同样,满足约束规范但不满足库恩—塔克条件的边界点,作为最优解完全可以排除。如果某点 $x^0$ 的确满足必要条件,但我们不能作出 $x^0$ 可构成一个 739 最优解的结论。因为某些经检验不合格的候选点完全可以通过筛选机制的检验,就像最简单的最优化问题中拐点可满足一阶条件 $dy/dx=0$ 一样。换言之,如果把必要条件比作鱼网,那么,我们捕到的可能是真的最优解,也可能是假的最优解。

有了充分条件,情况就不同了。因为如果点 $\bar{x}$ 满足极大值的充分条件,那么该点必使目标函数达到极大值。在这种意义上,充分条件提供了一类更明确地检验方法。但是它们也有自己的缺点,即充分条件本身可能不是必要条件,因而真正的最优解仍然可以不满足这个充分条件。换句话说,如果把充分条件视为鱼网,我们可能捕不到真正的最优解。

当然,最令人满意的情况是拥有充分必要条件。利用这个条

件我们能得到所有最优解，也不会冒把经检验不合格的候选点作为最优解的危险。

在这一节中，我们将证明，在某些情况下，可以将库恩—塔克条件视为极值的充分条件，甚至也可以将其视为充要条件。

### 库恩—塔克充分性定理

在古典最优化问题中，极大值和极小值的充分条件传统上是按照二阶导数或微分的符号来表达的。但正如我们已在 11.5 节中已证明的那样，这些二阶条件总是与目标函数的凹凸性有密切关系。这里，在非线性规划中，充分条件的表述也包含凹性和凸性的概念，只是现在这些概念不仅应用到目标函数，而且还要应用到约束函数 $g^i(x)$。

对于极大化问题，库恩和塔克给出下面的充分条件（充分性定理）

给定非线性规划：

Max. $\quad \pi = f(x)$

s.t. $\quad g^i(x) \leqslant r_i \qquad (i = 1, 2, \cdots, m)$

且 $\quad x \geqslant 0$

如果满足下列诸条件：

($a$) 目标函数 $f(x)$ 在非负正交分划体中可微，且为凹函数；

740 ($b$) 每个约束函数 $g^i(x)$ 在非负正交分划体中可微，且为凸函数；

($c$) 点 $\bar{x}$ 满足库恩—塔克极大化条件。

那末，$\bar{x}$ 为 $\pi = f(x)$ 的整体极大值点。

注意，在这个定理中，没有一处提到约束规范。这是因为在条

件$(c)$中我们已假设库恩—塔克条件在$\bar{x}$处满足，因此约束规范的问题便不需再提及了。

依此情况，上述定理表明条件$(a)$、$(b)$和$(c)$是使$\bar{x}$成为最优解的充分条件。但从不同角度看时，我们也可把它解释为：如果给定条件$(a)$和$(b)$，那末库恩—塔克极大化条件就是极大值的充分条件。从上一节我们知道，库恩—塔克条件本身虽然不是必要条件，但当满足约束规范时，就成为必要条件了。把这个认识与充分性定理结合起来，现在可表述为：如果满足约束规范，且实现条件$(a)$和$(b)$，那么库恩—塔克极大化条件就是极大化的**充分必要**条件。例如，当所有约束都是线性不等式，且充分满足约束规范时，就属于这种情况。后面，我们将介绍另外一些情况，即使函数$g(x)$不全是线性的，也能保证满足约束规范。

在上面充分性定理中讨论的极大化问题，通常称为**凹规划**。这个名称的产生是因为库恩和塔克在每个约束条件中用不等式$\geqslant$代替不等式$\leqslant$的缘故。因而条件$(b)$要求所有的函数$g^i(x)$像函数$f(x)$一样，都是**凹函数**，但为了与前面讨论的线性规划一致，我们在表述上已进行了修改。虽然形式不同，实质上两者是等价的。

如上所述，充分性定理仅对极大化问题进行了讨论。但对**极小化**问题同样也是适用的。除了定理适当变化以反应问题本身的反向变化以外，我们只需在条件$(a)$和$(b)$中把**凹**和**凸**两字互换，在条件$(c)$中使用库恩—塔克**极小化**条件即可（见练习 21.4 第 1 题）。

因为库恩—塔克充分性定理的证明比较容易，所以我们下面重述此定理的证明。

### 充分性定理的证明

对于极大化问题,拉格朗日函数可表示为

$$Z = f(x) + \sum_{i=1}^{m} \bar{y}_i [r_i - g^i(x)] \qquad [参见(21.17)]$$

$$(21.24)$$

741 此处我们已赋予拉格朗日乘数具体值 $\bar{y}_i$,从而使 $Z$ 单独是变量 $x_j$ 的函数。与充分性定理的条件($a$)和($b$)一致。我们假设 $f(x)$是凹函数,每个 $g^i(x)$是凸函数,因而使每个 $-g^i(x)$为凹函数。则因为 $Z$ 是凹函数的和,所以必然也是 $x$ 的凹函数。因此,按照(21.24),得到

$$Z(x) \leqslant Z(\bar{x}) + \sum_{j=1}^{n} \frac{\partial Z}{\partial \bar{x}_j}(x_j - \bar{x}_j) \qquad (21.25)$$

其中 $\bar{x}$ 是定义域中某个确定的点,且$\partial Z / \partial \bar{x}_j$ 是在 $\bar{x}$ 处计值的偏导数$\partial Z / \partial x_j$,特别地,我们确定选择变量和拉格朗日乘数的值分别为 $\bar{x}$ 和 $\bar{y}$,使之满足库恩—塔克极大化条件,并与充分性定理条件($c$)一致。那么可以证明(21.25)中表达式是非正的。由于去掉它之后,并不会使不等式反号,从而使我们可以推断出 $Z(x) \leqslant Z(\bar{x})$。

为了证明这一点,把$\sum$表达式分解为两项:

$$T_1 = \sum_{j=1}^{n} \frac{\partial Z}{\partial \bar{x}_j} x_j \qquad 和 \qquad T_2 = -\sum_{j=1}^{n} \frac{\partial Z}{\partial \bar{x}_j} \bar{x}_j$$

利用在点 $\bar{x}$ 处的互补松弛,$T_2$ 必变成零,至于 $T_1$,它含有 $x_j$ 而非 $\bar{x}_j$,因而互补松弛性不适用。我们只能肯定,对每个 $j$ 有$\partial Z / \partial \bar{x}_j \leqslant 0$(边际条件)和 $x_j \geqslant 0$(模型设定),于是 $T_1$ 为非正。把这两项

相加,得出(21.25)中的$\sum$表达式为非正。因此,我们可以得出结论

$$Z(x) \leqslant Z(\bar{x}) \tag{21.26}$$

或根据(21.24),有

$$f(x)+\sum_{i=1}^{m}\bar{y}_i[r_i-g^i(x)] \leqslant f(\bar{x})+\sum_{i=1}^{m}\bar{y}_i[r_i-g^i(\bar{x})] \tag{21.26'}$$

如果我们能证明删掉(21.26′)中两个$\sum$表达式不会使不等式反号,那么就能得出结论 $f(x)\leqslant f(\bar{x})$,这就证实了 $\bar{x}$ 使目标函数 $f(x)$ 达到极大的值的论断。而事实上确实能够做到这一点。$\sum$表达式的左边必然为非负,因为对每个 $i$,有 $\bar{y}_i\geqslant 0$(非负)和 $r_i-g^i(x)\geqslant 0$(约束条件的设定)。反之,$\sum$表达式的右边必然为零,因为$[r_i-g^i(\bar{x})\geqslant 0]$只不过$\partial Z/\partial\bar{y}_i$,因而互补松弛是适用的。因此 $f(x)\leqslant f(\bar{x})$,$\bar{x}$ 的确是最优解。

极大值$\bar{\pi}=f(\bar{x})$是**整体**极大值。证明这一结论的一种方法是证明不等式 $f(x)\leqslant f(\bar{x})$ 的正确性并不依赖于 $x$ 位于 $\bar{x}$ 的任意限定邻域。更正式地,我们可借助 19.3 节介绍的整体性定理。假设目标函数为可微的凹函数,至于约束条件,因每个 $g^i(x)$为凸函数,故(11.27)中的结果(当适当地推广到 $n$ 维情况后)可表明集合 742

$$S_i^{\leqslant}\equiv\{x \mid g^i(x)\leqslant r_i\}$$

一定是凸集。事实上,它还是一个闭凸集合,而且可行区域就是闭凸集合 $S_i^{\leqslant}$,$(i=1,2,\cdots,m)$的交,因而可行集也是一个闭凸集合。因此,能应用整体性定理,任意局部极大值必是整体极大值。注意,如果目标函数 $f(x)$是**严格**凹函数,那么整体极大值是唯一

的。

上面，从以下假定条件出发：($a$)函数 $f(x)$可微且为凹函数；($b$)每个函数 $g^i(x)$可微且为凸函数；($c$)点 $\bar{x}$满足库恩—塔克极大化条件，我们推出 $\bar{x}$ 的确使$\pi$达到极大值。因此充分性定理获证。

## 鞍　点

(21.26)中的结果(是在具体选择 $y$ 的基础上推导出来的)表明，给定 $\bar{y}$，在所有可接受的 $x$ 中，$\bar{x}$ 使拉格朗日函数 $Z$ 达到极大。实际上，已知 $\bar{x}$，在 $y$ 所容许值之中，$\bar{y}$ 使 $Z$ 达到极小也同样是成立的。即，事实上我们得到：

$$Z(x,\bar{y}) \leqslant Z(\bar{x},\bar{y}) \leqslant Z(\bar{x},y) \tag{21.27}$$

因为这种情况类似于图 11.3$a$ 中所描述的情况，故点($\bar{x}$，$\bar{y}$)称为鞍点，$Z(\bar{x},\bar{y})$称为拉格朗日函数的鞍值。

也许读者能够回忆起来，(21.27)的第一个不等式只不过是在明确提到$\bar{y}$为给定的前提下，根据给定 $\bar{y}$，$Z$ 是以 $x$ 为变量的凹函数这一事实，对(21.26)的重述。据此，我们也可以证明，(21.27)的第二个不等式是与以下事实相联系的，给定 $\bar{x}$，$Z$ 是变量$y$ 的凸函数。在拉格朗日函数中令 $x=\bar{x}$，得到：

$$Z = f(\bar{x}) + \sum_{i=1}^{m} y_i[r_i - g^i(\bar{x})]$$

因 $r_i$ 和$g^i(\bar{x})$表示常数，显然 $Z$ 是变量 $y_i$ 的线性函数，也是凸函数。参考(21.24′)，可知 $Z$ 对 $y_i$ 的凸性具有如下含义：

$$Z(y) \geqslant Z(\bar{y}) + \sum_{i=1}^{m} \frac{\partial Z}{\partial \bar{y}_i}(y_i - \bar{y}_i) \qquad [\text{参见 } (21.25)] \tag{21.28}$$

其中我们可将 $\bar{y}$ 视为满足库恩—塔克极大化条件的 $y$ 值。则用分析(21.25)同样的方法,我们可证明(21.28)中的$\sum$表达式为非负,因而删去它并不影响不等式的正确性。从而,得到 $Z(\bar{y})\leqslant$ 743 $Z(y)$,或者更彻底地,$Z(\bar{x},\bar{y})\leqslant Z(\bar{x},y)$。根据上面的讨论,我们可把这样的点称为鞍点$(\bar{x},\bar{x})$:在该点,给定函数 $Z(x,y)$对一组自变量(此处为 $x$)为凹函数,但对另一组自变量(此处为 $y$)为凸函数。

将鞍点的概念应用于线性规划时,它显得特别重要,因为它可直接引出对偶的概念。一般而言,一对原规划和对偶规划可用向量符号写成:

| | | | |
|---|---|---|---|
| Max. | $\pi=c'x$ | Min. | $\pi^*=r'y$ |
| s.t. | $Ax\leqslant r$ | s.t. | $A'y\geqslant c$ |
| 且 | $x\geqslant 0$ | 且 | $y\geqslant 0$ |

两种规划的拉格朗日函数(21.17)表现为如下形式:

$$Z=c'x+y'(r-Ax)=c'x+y'r-y'Ax$$

[原函数]

和 $$Z'=r'y+x'(c-A'y)=r'y+x'c-x'A'y$$

[对偶函数]

注意 $Z$ 和$Z'$互为转置,因为 $Z$ 和$Z'$都是 $1\times 1$ 的矩阵,显然,它们表示相等的标量。因而,我们使拉格朗日函数对变量 $x_j$ 取极大值时(对偶规划),实际上同时使拉格朗日函数对变量 $x$ 取极大值(原规划)。在这两个规划中得到最优解 $\bar{x}$ 和 $\bar{y}$,那时它们必然相应地满足不等式(21.27)。这就证明了对偶概念和鞍点概念实际上是相同的。

根据互补松弛性,$Z$ 的$y'(r-Ax)$分量在最优解的情况下必

然为零[参阅(21.14)的最后一个方程]。类似地,$Z'$中的 $x'(c-A'y)$部分必然也为零。这些事实揭示了 20.1 节对偶定理的实质。更进一步,通过分别从 $Z$ 和 $Z'$中减去为零的部分,我们求得在最优解的情况下,有 $\bar{Z}=c'\bar{x}=\bar{\pi}$ 和 $\bar{Z}'=r'\bar{y}=\bar{\pi}^*$。则 $\bar{Z}=\bar{Z}'$ 的事实就意味着$\bar{\pi}=\bar{\pi}^*$,这就是对偶定理 I 的实质。这样从非线性规划角度考察这个问题时,线性规划的对偶特性便更清楚地表现了出来了。

## 练习 21.4

---

1 已知 Min. $C=F(x)$

s.t. $G^i(x)\geqslant r_i(i=1,2,\cdots,m)$

且 $x\geqslant 0$

($a$) 把它变为极大化问题。

($b$) 库恩—塔克充分性定理的函数 $f$ 和 $g^i$ 在这个问题中相当于什么?

($c$) 为了能在这里应用极大的充分条件,对函数 $F$ 和 $G^i$ 应施加什么样的凹凸性条件?

744

($d$) 根据上述内容,我们应如何表述极小值的库恩—塔克充分条件?

2 证明:当上一问题中的极小化规划满足极小值的充分条件时,可应用整体性定理,因而获得的极小值也是整体极小值。

3 在(21.27)中 $Z(\bar{x},\bar{y})$是唯一的鞍点值吗?如何修改(21.27)以描述唯一的鞍点值?

4 库恩—塔克充分性定理可应用于下列问题吗?

($a$) Max. $\pi=x_1$

s.t. $x_1^2+x_2^2\leqslant 1$

且 $x_1,x_2\geqslant 0$

($b$) Min. $C=(x_1-3)^2+(x_2-4)^2$

s.t. $\quad x_1 + x_2 \geqslant 4$

且 $\quad x_1, x_2 \geqslant 0$

(c) Min. $\quad C = 2x_1 + x_2$

s.t. $\quad x_1^2 - 4x_1 + x_2 \geqslant 0$

且 $\quad x_1, x_2 \geqslant 0$

---

## 21.5 阿罗—恩索文充分性定理:拟凹规划

为了应用库恩—塔克充分性定理,必须满足某些凹凸性规定。这是非常严格的要求。这里要介绍的另一个充分性定理——阿罗—恩索文充分性定理[①]——则弱化了这些要求。它只要求目标函数和约束函数具有拟凹性和拟凸性。由于对凸凹性的要求减弱了,充分条件的应用范围相应地就扩大了。

在阿罗—恩索文的原始论文中,具有极大化问题和形式为≥的约束不等式,函数 $f(x)$ 和 $g^i(x)$ 一律都是拟凹函数,以便使定理能够应用。这就产生了拟凹规划这个名称。但这里的讨论中,我们在极大化问题的约束中仍使用≤的不等式,在极小化问题的约束中,仍使用≥的不等式。

### 阿罗—恩索文充分性定理

745

定理叙述如下:

已知非线性规划:

---

① K.J.阿罗和A.C.恩索文:"拟凹规划"载《经济计量学》1961年,第779—800页。

$$\begin{aligned} &\text{Max.} \quad \pi = f(x) \\ &\text{s.t.} \quad g^i(x) \leqslant r_i \qquad (i = 1, 2, \cdots, m) \\ &\text{且} \quad x \geqslant 0 \end{aligned}$$

如果满足下列条件：

($a$) 在非负正交分划体中目标函数 $f(x)$可微且是拟凹函数；

($b$) 在非负正交分划体中每个约束函数 $g^i(x)$可微，且是拟凸函数；

($c$) 点 $\bar{x}$ 满足库恩—塔克极大条件

($d$) 满足下列诸条件中任意一个：

($d-i$) 至少对某个变量 $x_j$ 有 $f_j(\bar{x})>0$。

($d-ii$) 对某个可取正值而不违背约束的变量 $x_j$ 有 $f_j(\bar{x})>0$。

($d-iii$) $n$ 个导数有 $f_j(\bar{x})$不全为零，函数 $f(x)$在 $\bar{x}$ 的邻域内二阶可微[即在 $\bar{x}$处，$f(x)$的所有二阶偏导数都存在。]

($d-iv$) 函数 $f(x)$为凹函数。

那么 $\bar{x}$ 给出函数$\pi=f(x)$的整体极大值。

因为这个定理的证明有些复杂，在这里我们就省略了。但希望读者对这个定理的几个重要特征要给予关注。首先，尽管阿罗和恩索文成功地把凹凸性的规定相应地减弱到拟凹拟凸的要求，但他们发现附加新要求($d$)也是必要的。不过要注意，为了组成一组完全的充分条件，只要求($d$)项中四个可供选择的条件中一个成立即可。因此，事实上，上面定理含有四组不同的极大充分条件。在($d-iv$)的情况下，$f(x)$为凹函数，看来好像阿罗—恩索文充分性定理与库恩—塔克充分性定理是一致的。但这是不正确

的，因为阿罗和恩索文只要求约束函数 $g^i(x)$ 是拟凹函数，所以其充分条件仍然是较弱的。

如前所述，定理是把条件 $(a)$ 到条件 $(d)$ 并在一起作为一组充分条件的。但也可解释为：当满足条件 $(a)$、$(b)$ 和 $(d)$ 时，那么库恩—塔克极大化条件变成极大化的充分条件。而且，如果又满足约束规范，那么库恩—塔克条件变成极大化的充分必要条件。 746

像库恩—塔克定理一样，也很容易使阿罗—恩索文定理适合于极小化的结构。除了需要调换最优化的方向这个明显变化以外，我们只要在条件 $(a)$ 和 $(b)$ 中互换拟凹和拟凸这两个词，用极小化条件取代库恩—塔克极大化条件；把 $(d-i)$ 和 $(d-ii)$ 中的不等式反号；且在 $(d-iv)$ 中把凹字变为凸字。

## 约束规范的检验方法

前面早已提到过，如果所有约束函数是线性的，那么约束规范就会得到满足。如果函数 $g^i(x)$ 是非线性的，那么在确定是否满足约束规范方面，阿罗和恩索文提出的下列检验法也是有用的：

对极大化问题，如果

$(a)$ 每个约束函数 $g^i(x)$ 可微且为拟凸函数；

$(b)$ 在非负正交分划体存在一点 $x^0$，使在 $x^0$ 处满足作为严格不等式的所有约束条件。

$(c)$ 下列诸条件中一个成立：

$(c-i)$ 每个函数 $g^i(x)$ 为凸函数。

$(c-ii)$ 每个 $g^i(x)$ 的偏导数在可行区域中的每点 $x$ 计值时，不全为零。那末，约束规范得到满足。

而且，我们也很容易使这个检验法适用于极小化问题。为此，只要

在条件($a$)中把拟凸这个词变为拟凹;在($c-i$)中把凸这个字改为凹。这个检验法的应用将在后面举例说明。

**练习 21.5**

---

1　当把阿罗—恩索文充分性定理和约束规范应用于极小化问题时,证明课文中对这定理和规范所进行的变化是正确的。

2　约束规范检验法在本质上是一组必要条件呢? 还是一组充分条件呢?

3　下列哪些函数作为适合于应用阿罗—恩索文充分性定理的极大化问题的目标函数,在数学上是可以接受的?

($a$) $f(x)=x^3-2x$　　($b$) $f(x_1,x_2)=6x_1-9x_2$

($c$) $f(x_1,x_2)=x_2-\ln x_1$

(提示:参见练习 12.4-4)

4　已知极大化问题的约束条件是:

747　($a$) $x_1^2+(x_2-5)^2\leqslant 4$　和　$5x_1+x_2\leqslant 10$

($b$) $x_1+x_2\leqslant 8$　和　$-x_1x_2\leqslant -8$(注意: $-x_1x_2$不是凸的)

它们满足约束规范吗?

---

## 21.6　经济应用

现在我们来介绍这两个充分性定理在经济上的某些简单应用。

### 最大效用问题再研究

在前面 2.4 节中,我们遇到过消费理论问题。它是用下列形

式表述的:在满足 $P_1x_1+P_2x_2=B$ 的情况下,使 $U=U(x_1,x_2)$ 极大化。约束条件表现为一个方程,选择变量没有具体限制取非负值。现在我们可用比较结合实际的形式重新表述此问题。考虑 $n$ 个变量时,可写成非线性规划形式:

$$\begin{aligned} &\text{Max.} \quad && U=U(x_1,\cdots,x_n) \\ &\text{s.t.} \quad && P_1x_1+\cdots+P_nx_n\leqslant B \\ &\text{且} \quad && x_1,\cdots,x_n\geqslant 0 \end{aligned} \tag{21.29}$$

其中所有价格都是外生的。在这个新的公式中,消费者可以选择少于预算数量 $B$ 的支出。另外,在这个问题中,任何商品的购买量不能为负这一事实,也得到了明确的陈述。

因(仅有的)约束函数 $g^1(x)=P_1x_1+\cdots+P_nx_n$ 是线性的,所以满足约束规范,且如果两个充分性定理中所提到的其他条件都满足的话,那么,库恩—塔克极大条件就是充分必要条件。为了应用库恩—塔克定理,我们要求 $U(x)$ 可微,且是凹函数;$g^1(x)$ 可微,且为凸函数。因为 $g^1(x)$ 为线性函数,当然满足所要求的条件,但 $U(x)$ 的条件必须由模型明确规定。但在检验阿罗—恩索文定理时,我们要注意,如果条件 $(d)$ 正好满足,仅需规定 $U(x)$ 是可微函数和拟凹函数即可——相当不严格的要求。例如,如果有某种商品 $j$ 在消费者预算范围之内,而且总是产生正的边际效用(不饱和),那么,当然有 $U_j(\bar{x})>0$,且满足条件 $(d-ii)$。自然,还有其他方法满足条件 $(d)$,使库恩—塔克极大条件成为充分必要条件。

拟凹性的要求完全与前面在12.5节中所讨论的古典消费理论一致。大家回顾一下,在那里,两种商品模型有一个等式约束,规定函数 $U(x)$ 是严格拟凹函数。严格性假定的目的是为了

748

在 $x_1x_2U$ 空间中铃形状的效用曲面上排除出现水平台的情况，因而每个等值集合（这里是无差别集合）以一个细曲线而不是宽带的形式出现。因为只有在那种情况下，古典意义上的切点最优解才有意义。当预算约束条件为不等式时，情况就不同了。因为我们寻找的不再是切点，而是可行区域内的最优点。现在在效用曲面上有水平台是允许的，这样就不需要规定严格拟凹了。

在 12.5 节中，也假定函数 $U(x)$ 是 $x_1$ 和 $x_2$ 的递增函数。在经济上，这表示没有达到饱和。在图形上，这是把效用曲面仅限定在铃形的上升部分，因而如图 21.7 所示，无差异曲线通常向下倾斜，且为凸型曲线，而不是圆形或卵形的自闭曲线。这样假设的理由在古典结构切点解的特性中可以找到。在此情况下，虽然像 $H$

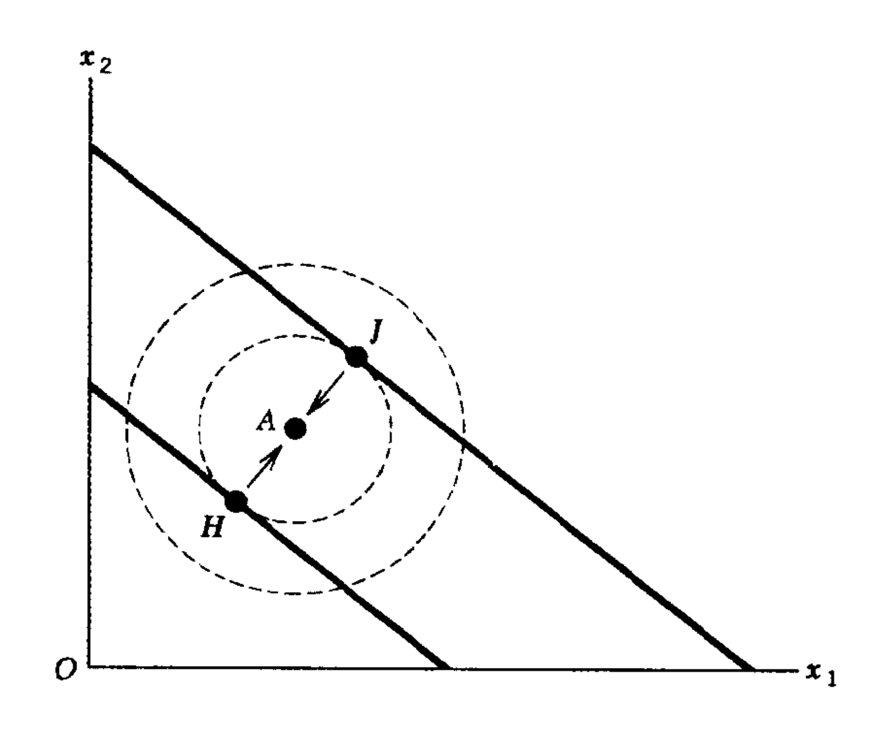



图 21.7

这样的切点确实表示最好的选择（因为在基平面中，经过点 $H$ 向点 $A$ 移动，上升到效用曲面的峰点，而预算约束使这种情况不可

能发生)。但离峰点较远一些的切点,例如点 $J$ 显然不是最优解,被排除在外。转到非线性规划的结构时,我们看到非饱和的假设也是有用的——用来满足条件 $(d-ii)$。但是因为在阿罗—恩索文充分性定理中的条件 $(d)$ 可用其他方法加以满足,故完全没有必要假定 $U(x)$ 是递增的。

本问题的库恩—塔克条件容易写出和解释,因此留给读者作为练习。

## 最小成本组合再研究



假设厂商的生产函数是一种科布—道格拉斯函数:$Q=K^{\alpha}L^{\beta}$,式中 $\alpha>0$, $\beta<1$,再令投入价格 $P_k>0, P_L>0$(资本服务的价格和劳动服务的价格)是由外因因素确定的,那么最小成本组合的问题可用下列非线性规划来表达:

$$
\begin{array}{lll}
\text{Min.} & C=P_K K+P_L L & \\
\text{s.t.} & K^{\alpha}L^{\beta}\geqslant Q_0 \qquad (Q_0>0) & \qquad (21.30)\\
\text{且} & K, L\geqslant 0 &
\end{array}
$$

此公式有别于12.7节的公式。它允许厂商生产多于需要的数量 $Q_0$,并明确承认使用的投入数量不能为负这一事实。事实上,给定 $Q_0$ 为正,在科布—道格拉斯函数生产函数情况下,$K$ 和 $L$ 一定均为正。

因为目标函数是线性的,它们自然是凸函数,也是拟凸函数。(唯一的)约束函数 $K^{\alpha}L^{\beta}$,对正的 $K$ 和 $L$ 是拟凹函数,与 $\alpha+\beta\gtreqless 1$ 无关。即与规模收益是递增、不变还是递减无关[见12.4节例5和(12.52)的讨论]。但是,如果 $\alpha+\beta>1$(规模收益递增),那么,

这个函数不是凹函数。在后一种情况下，要求每个单位产出的连续增长大于两种投入比例科布—道格拉斯函数的增长。在图形上，当我们偏离原点沿 $KL$ 平面中的任意已知射线移动时，产出单位增量所形成的一系列等产量曲线越来越密了。因此，这样的射线投影到生产曲面得到一曲线，它表明以越来越快的速度上升，因而曲面不可能是凹的。

由于这种情况，当规模收益递增不断递增时，库恩—塔克定理就不适用了。反之，阿罗—恩索文定理即使在这种情况下仍能应用。对于极小化问题，阿罗—恩索文定理要求目标函数为拟凸函数，约束为拟凹函数。这些要求在现在的问题中是得到满足的。此外，条件($d-i$)也满足。因为在极小化问题中，对某个选择变量 $x_j$，假定我们有 $f_j(\bar{x})>0$，而且我们确实有$\partial C/\partial K=P_K>0$ 和 $\partial C/\partial L=P_L>0$。

深入一步，我们还可断言，约束函数 $g^1=K^\alpha L^\beta$ 经得起约束规范的检验。首先，$g^1$ 是拟凹函数，第二，一定存在非负的有序偶($K,L$)满足严格不等式的约束条件。第三，因可行区域中每点具有 $K>0$ 和 $L>0$ 的特征（否则 $Q_0$ 不能为正），偏导数$\partial g^1/\partial K=\alpha K^\alpha L^{\beta-1}$和$\partial g^1/\partial L=\beta K^\alpha L^{\beta-1}$在可行区域内处处为正，于是满足检验法的条件($c-ii$)。因此，由于满足约束规范，库恩—塔克极小化条件就可看作必要和充分条件。



## 销售最大化的厂商

标准的厂商分析，一般其具有利润最大目标。但若所研究的厂商是所有者和管理者分开的公司，则将管理者的目标确定为销

售(收益)最大化,也是很有道理的。[①] 总收益通常被视为产业内部厂商竞争地位的重要指标,而且销售额的增长常被视为管理成功的标志。可以想象,管理者的报酬也直接依赖于这个特定的绩效指标。所以,销售量大看来是公司机构可供选择的一个目标,当然,为了避免股东可能的不满意,管理者还应努力做到利润水平不低于某个规定的最小值,比如$\pi_0$。这个$\pi_0$低于与 $MR = MC$ 条件相联系的最大利润水平。

如果是这样,管理者的问题就是使 $R = R(Q)$最大化,满足约束$\pi = R(Q) - C(Q) \geqslant \pi_0$,或者

$$\begin{aligned} &\text{Max.} && R = R(Q) \\ &\text{s.t.} && C(Q) - R(Q) \leqslant -\pi_0 \qquad (\pi_0 > 0) \\ &\text{且} && Q \geqslant 0 \end{aligned}$$

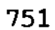
只要 $R(Q)$可微且是凹函数,$C(Q)$可微且是凸函数[这意味着约束函数 $C(Q)—R(Q)$可微且是凸函数],就可应用库恩—塔克充分性定理。注意到这一点是重要的,即在这种情况下,我们不能寄希望于通过将凹凸性假定放松至拟凹的 $R(Q)$和拟凸的 $C(Q)$,来获得更大的一般性。因而,阿罗—恩索文充分性不能自然得到应用,当然,除非约束函数通过直接的新的假设而得到了拟凸性。

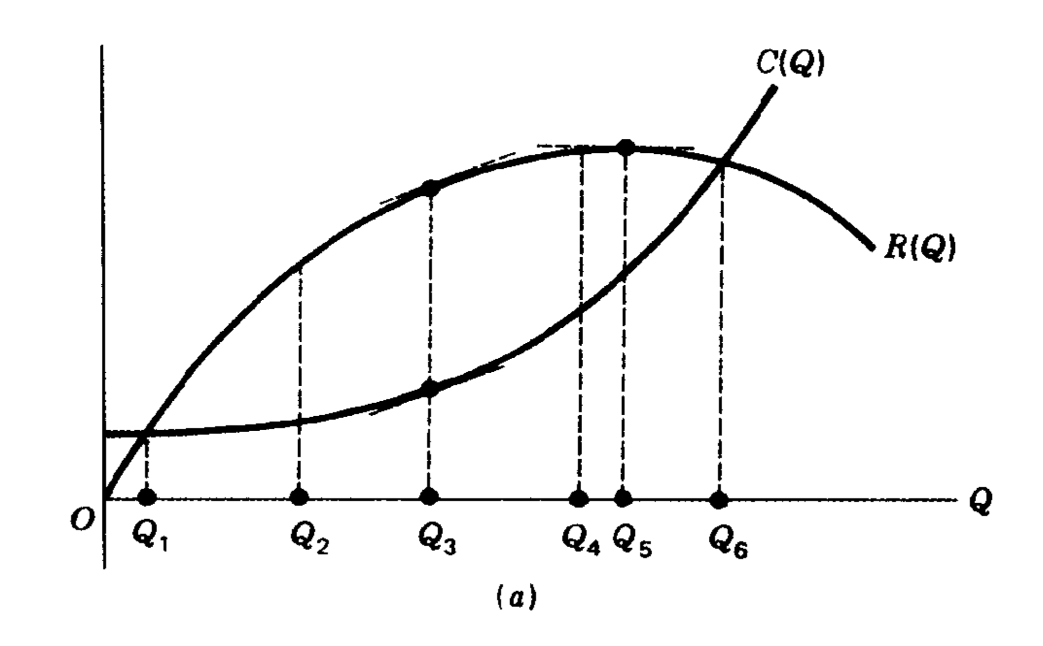
在图 21.8 中,我们描绘了凹的 $R(Q)$函数和凸的 $C(Q)$函数的情况。图 $a$ 中的两条曲线是在$R(0)=0$ 和 $C(Q)>0$ 的假设下绘出的。图 $b$ 中的曲线(是对水平轴的镜像)表示约束函数,是利润曲线的负值。尽管在开放区间($Q_1, Q_6$)的任意产出水平能够

---

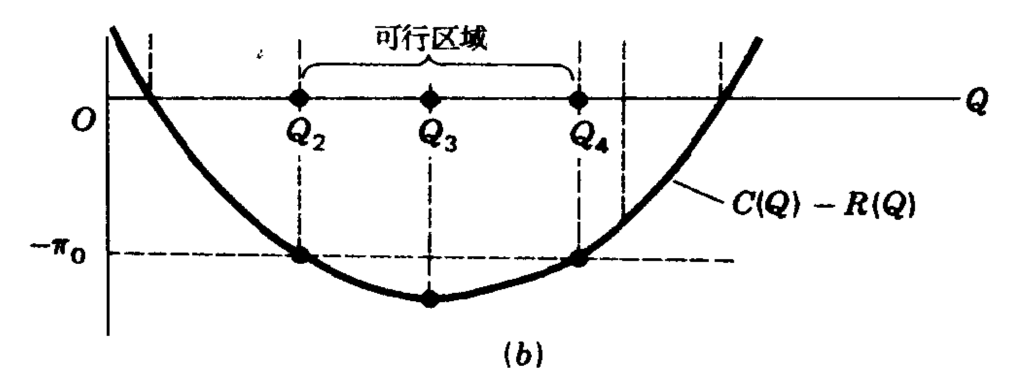
① W.J. 鲍莫尔:《论寡头垄断理论》,收载于《经济计量学》,1958 年 8 月,第 187—198 页,也可参看他著的《商业行为,价值和增长》,修订版,Harcourt, Brace & World Inc., New Jork, 1967 年。

产生正的利润,但只有在封闭区间$[Q_2,Q_4]$的 $Q$ 水平才会产生不小于$\pi_0$的利润,并满足利润约束。因此,我们可以确定$[Q_2, Q_4]$可行区域。正如可以预期的那样,可行区域包括 $Q_3$,这个产出水平使利润最大化。






(a)




(b)

图 21.8

这个问题的拉格朗日函数为:

$$Z = R(Q) + y[-\pi_0 - C(Q) + R(Q)]$$

库恩—塔克条件是由边际条件

$$\frac{\partial Z}{\partial Q} = R'(Q) - yC'(Q) + yR'(Q) \leqslant 0 \qquad (21.31)$$

$$\frac{\partial Z}{\partial y}=-\pi_0-C(Q)+R(Q)\geqslant 0$$

加上非负性和互补松弛条件所组成。因为 $R(0)=0$ 和 $C(0)>0$，所以，零产出得到 $\partial Z/\partial y=-\pi_0-C(0)<0$，这就违背第二个边际条件。因此我们必取 $Q>0$（这个要求与零产出水平位于可行区域之外这一事实是完全一致的）。根据互补松弛性，$Q$ 为正意味着 $\partial Z/\partial Q=0$。即(21.31)中的第一个弱不等式必然被作为一个等式而满足。解后一方程，我们得到销售最大化产出的法则　752

$$R'(Q)=\frac{y}{1+y}C'(Q) \tag{21.32}$$

在这个结果中，$y$ 的值或者为零，或者为正。若 $y=0$，此法则简化为 $R'(Q)=0$，厂商会设法把其产出推到边际收益为零的水平，即图 21.8 中的 $Q_5$。从形式上讲，这是最典型的销售最大化。因为在这种情况下，厂商已达到总收益的顶峰。但在我们的假设下这种极端的行为是不可行的，因为 $Q_5$ 也位于可行区域之外。相应地，我们必须取 $y>0$。根据互补松弛性，这意味着 $\partial Z/\partial y=0$，而这又意味着利润约束被作为等式满足，厂商力图恰好获得所要求的最小利润 $\pi_0$。

具有正的 $y$，销售额最大化产出法则(21.32)表明

$$R'(Q)<C'(Q) \qquad \left[\text{因为}\ \frac{y}{1+y}<1\right]$$

在一般情况下，这会比利润最大化法则 $R'(Q)=C'(Q)$ 获得更大的产出。利润最大化的产出水平为 $Q_3$。满足法则(21.32)并产生恰好是所需的最小利润量的产出水平是 $Q_4$，它确实大于 $Q_3$。

## 利用库恩—塔克条件解非线性规划

迄今为止，我们对库恩—塔克条件的讨论主要集中于其（必要

条件和充分条件方面的)分析作用。如果这些条件恰巧是充分必要条件,[①] 且选择变量个数适当小,那么它们在计算方面也能发挥作用。我们用厂商销售额最大的数字例子来介绍这种方法。

假设厂商的收益函数和成本函数为

$$R = 32Q - Q^2 \quad [凹] \quad 和 \quad C = Q^2 + 8Q + 4 \quad [凸]$$

且最小利润为$\pi_0 = 18$。在这些情况下,库恩—塔克条件实际上是充分必要条件。根据(21.31),两个边际条件取如下形式:

$$\frac{\partial Z}{\partial Q} = 32 - 2Q - y(4Q - 24) \leqslant 0$$

$$\frac{\partial Z}{\partial y} = -2Q^2 + 24Q - 22 \geqslant 0$$

753 $Q$ 的值或为正,或等于零。先试 $Q = 0$,我们马上在第二个边际条件中遇到矛盾。这样,我们必假设 $Q > 0$。如果是这样,就可由互补松弛推出$\partial Z / \partial Q = 0$,它给出含两个变量的方程。在这个方程中,$y$ 的值必为零,或为正。试探 $y = 0$,解出 $Q$,得到 $Q = 16$。虽然仅就第一个方程而论,$Q = 16$ 是允许的。但 $Q$ 的这个值表示$\partial Z / \partial y = -150$,与第二个边际条件(利润约束条件)相违背。于是,我们只好令 $y > 0$。由互补松弛得出$\partial Z / \partial y = 0$。解这个新方程,我们得到两个根 $Q_1 = 11$ 和 $Q_2 = 1$。但因只有第一个根与条件$\partial Z / \partial Q = 0$ 一致,所以销售最大化产出为 $\bar{Q} = 11$。反之,$R'(Q) = C'(Q)$法则却得到一个较低的数字 $Q = 6$,作为利润最大的产出。

从上面显然可以看出,这种计算方法本质上是试错法。基本

---

① 若它们只是充分条件而非必要条件,则最优解可能不满足库恩—塔克条件,从而使后者作为一个"渔网",在捕捉这样的解方面是无效的。

思想是，首先对每个选择变量试以零值。令变量等于零时，由于使某些项消失，便可简化边际条件。然后，如果发现拉格朗日乘数的适当非负数满足所有的边际条件，那么这个零解将是最优的。另一方面，如果零解与某些不等式相违背，那么我们必须令一个或一个以上选择变量为正。对每个正的选择变量，由互补松弛性可以将不等式的边际条件转化为严格的等式。适当解出这样的方程后，或者得到一个解，或者导致矛盾。后一情况就迫使我们尝试另一个解值。如果解存在，用这种尝试法，最终总能发现它。

注意，在含两个变量的问题中，我们必须试验的选择变量可能的符号组合不少于四($2^2$)个：(0,0)(0,+)(+,0)和(+,+)。三个变量的问题有八个($2^3$)可能性需要检验。随着选择变量个数增多，复杂性也随之剧增，因此，对于高维非线性规划问题，这种方法是不合适的。

## 练习 21.6

---

1　写出问题(21.29)的库恩—塔克条件。

2　写出问题(21.30)的库恩—塔克条件。

3　假设销售额最大的厂商，其总收益 $R$ 依赖于产出 $Q$ 和广告费用 $A$，其总成本由生产成本 $C$ 和广告成本 $A$ 所组成。

($a$) 重新构建课文中给出的非线性规划。

($b$) 对函数 $R$ 和 $C$ 应加什么限制，能使库恩—塔克充分性定理可以应用？

754

($c$) 施加什么约束条件，能使阿罗—恩索文充分性定理可以应用？

4　Min.　$C = x_1^2 + x_2^2$

s.t.　$x_1 + x_2 \geqslant 2$

且 $x_1, x_2 \geqslant 0$

($a$)库恩—塔克充分性定理可以应用这个问题吗？库恩—塔克极小值条件是充分必要条件吗？

($b$)写出库恩—塔克条件，并利用试错法找出最优解。值 $\bar{x}_1$ 和 $\bar{x}_2$ 等于多少？

5 ($a$)构建一个线性规划来回答下述问题：什么是从原点(0,0)到通过点(0,2)和(2,0)的直线上或其上方的任意点的最短距离(以 $d$ 表示)？(提示：利用勾股定理。)

($b$)这个规划问题与上题中给出的规划问题有什么不同？前一问题的解($\bar{x}_1, \bar{x}_2$)可以作为该问题的解吗？为什么？

($c$)最短距离 $\bar{d}$ 等于多少？

---

## 21.7 数学规划的局限性

因为数学规划能够处理不等式约束，所以，它极大地扩大了我们对最优化问题的讨论范围。除了可以明显地应用于工商管理中的实际问题以外，它也可以使经济学家从一个新的角度考察消费理论、生产理论和资源配置理论。

但是，同我们讨论的其他方法一样，数学规划也不是完美无缺的。首先，在上面的讨论中，我们总是假定选择变量是连续的。但实际上，常有一个或多个变量只能取整数值(例如，3.75 架飞机没有什么实际意义)。幸好**整数规划**(数学规划方法的一个分支)很好地解决了这个困难。这种规划只得到整数值。

一个更严重的缺陷(不仅数学规划具有这种缺陷，而且本书所考察的所有最优化框架均有这种缺陷)，是解的静态性质。在写出一个最优解，比如($\bar{x}_1, \cdots, \bar{x}_n$)时，我们用其表示在一组给定的环

境中，每个选择变量所能作出的最好选择。但是，因为每个 $\bar{x}_j$ 表示单个数值，因此，它只能与某一单个的时点，或问题所处的环境假定没有发生变化的某个时期相联系。在上述两种情况下，问题及其解都是静态的。与此相反，动态最优化问题需要求出在给定时期每个选择变量完整的最优时间路径，而不仅仅是一个最优值。但要解决这样的动态最优化问题，我们必须首先具备变分法、最优控制理论和动态规划方面的知识。由于在有限的篇幅内，清楚透彻地阐明这些专题几乎是不可能的，所以，最好是将其另文讨论。 755

我们就这样通过指出有关方法和分析的局限性而结束了本书这部分内容的讨论。这样做的目的，决不是对我们刚刚非常辛苦地介绍的数学方法感到怀疑，而是提醒读者不要将其视为无所不能的法宝。事实上，清楚地了解所学习的方法的局限性，是我们学习中非常关键的一部分。若非如此，我们就会变成方法的奴隶，而非方法的主人！

# 附录 I　希腊字母

| | | |
|---|---|---|
| Α | α | alpha |
| Β | β | beta |
| Γ | γ | gamma |
| Δ | δ | delta |
| Ε | ε | epsilon |
| Ζ | ζ | zeta |
| Η | η | eta |
| Θ | θ | theta |
| Ι | ι | iota |
| Κ | κ | kappa |
| Λ | λ | lambda |
| Μ | μ | mu |
| Ν | ν | nu |
| Ξ | ξ | xi |
| Ο | ο | omicron |
| Π | π | pi |
| Ρ | ρ | rho |
| Σ | σ | sigma |
| Τ | τ | tau |
| Υ | υ | upsilon |
| Φ | ϕ（或 φ） | phi |
| Χ | χ | chi |
| Ψ | ψ | psi |
| Ω | ω | omega |

# 附录 II　数学符号

## 1. 集合

757

| | |
|---|---|
| $a \in S$ | $a$ 属于集合 $S$ 中的元素 |
| $b \notin S$ | $b$ 不属于集合 $S$ 中的元素 |
| $S \subset T$ | 集合 $S$ 为集合 $T$ 的子集 |
| $T \supset S$ | 集合 $T$ 包括集合 $S$ |
| $A \cup B$ | 集合 $A$ 与集合 $B$ 的并集 |
| $A \cap B$ | 集合 $A$ 与集合 $B$ 的交集 |
| $\tilde{S}$ | 集合 $S$ 的补集 |
| $\{\ \}$ 或 $\phi$ | 零集(空集) |
| $\{a, b, c\}$ | 具有元素 $a$、$b$、$c$ 的集合 |
| $\{x \mid x$ 具有性质 $P\}$ | 具有性质 $P$ 的所有个体的集合 |
| $\min\{a, b, c\}$ | 特定集合的最小元素 |
| $R$ | 所有实数的集合 |
| $R^2$ | 2 维实空间 |
| $R^n$ | $n$ 维实空间 |
| $(x, y)$ | 有序偶 |
| $(x, y, z)$ | 有序三元组 |
| $(a, b)$ | 从 $a$ 到 $b$ 的开区间 |
| $[a, b]$ | 从 $a$ 到 $b$ 的闭区间 |

## 2. 矩阵与行列式

758

| | |
|---|---|
| $A'$ 或 $A^T$ | 矩阵 $A$ 的转置 |
| $A^{-1}$ | 矩阵 $A$ 的逆 |
| $\lvert A \rvert$ | 矩阵 $A$ 的行列式 |
| $\lvert J \rvert$ | 雅可比矩阵 |

| 符号 | 含义 |
|---|---|
| $\|H\|$ | 海塞行列式 |
| $\|\overline{H}\|$ | 海塞加边行列式 |
| $r(A)$ | 矩阵 $A$ 的秩 |
| 0 | 零矩阵 |
| $u \cdot v$ | 向量 $u$ 和 $v$ 的内积(点积) |
| $u'v$ | 两个向量的标量积 |

## 3. 微积分

| 符号 | 含义 |
|---|---|
| 已知 $y=f(x)$, | 单变量 $x$ 的函数 |
| $\lim\limits_{x\to\infty} f(x)$ | 当 $x$ 趋于无穷时,$f(x)$的极限 |
| $dy$ | $y$ 的一阶微分 |
| $d^2y$ | $y$ 的二阶微分 |
| $\frac{dy}{dx}$或$f'(x)$ | 函数 $y=f(x)$的一阶导数 |
| $\frac{dy}{dx}\|_{x=x_0}$或 $f'(x_0)$ | 在 $x=x_0$ 处的导数 |
| $\frac{d^2y}{dx^2}$或 $f''(x)$ | $y=f(x)$的二阶导数 |
| $\frac{d^ny}{dx^n}$或$f^{(n)}(x)$ | $y=f(x)$的 $n$ 阶导数 |
| $\int f(x)dx$ | $f(x)$的不定积分 |
| $\int_a^b f(x)dx$ | $f(x)$从 $x=a$ 到 $x=b$ 的定积分 |

已知 $y=f(x_1,x_2,...,x_n)$:

| 符号 | 含义 |
|---|---|
| $\frac{\partial y}{\partial x_i}$或$f_i$ | $f$ 对 $x_i$ 的偏导数 |
| $\frac{dy}{dx_i}$ | $f$ 对 $x_i$ 的全导数 |
| $\frac{\S y}{\S x_i}$ | f 对 $x_i$ 的偏全导数 |

## 4. 微分方程与差分方程

| 符号 | 含义 |
|---|---|
| $\dot{y}\equiv\frac{dy}{dt}$ | $y$ 对时间的导数 |

| | |
|---|---|
| $\Delta y_t$ | $y_t$ 的一阶差分 |
| $\Delta^2 y_t$ | $y_t$ 的二阶差分 |
| $y_p$ | 特别积分 |
| $y_c$ | 余函数 |

## 5. 其它符号

| | |
|---|---|
| $\sum_{i=1}^{n} x_i$ | 当 $i$ 由 1 至 $n$ 时，$x_i$ 的和 |
| $p \Rightarrow q$ | 仅当 $q$，则 $p$（$p$ 意味着 $q$） |
| $p \Leftarrow q$ | 若 $q$，则 $p$（$q$ 意味着 $p$） |
| $p \Leftrightarrow q$ | 当且仅当 $q$，则 $p$ |
| $iff$ | 当且仅当 |
| $\|m\|$ | $m$ 的绝对值 |
| $n!$ | $n$ 的阶乘 $\equiv n(n-1)(n-2)\ldots(3)(2)(1)$ |
| $\log_b x$ | 以 $b$ 为底 $x$ 的对数 |
| $\log_e x$ 或 $\ln x$ | $x$ 的自然对数（以 $e$ 为底） |
| $e$ | 自然对数和自然指数函数的底 |
| $\sin\theta$ | $\theta$ 的正弦函数 |
| $\cos\theta$ | $\theta$ 的余弦函数 |
| $R_n$ | 包含 $n$ 次多项式的泰勒级数的余项 |

# 附录 III 主要参考文献


Abadie, J. (ed.): *Nonlinear Programming*, North-Holland Publishing Company, Amsterdam, 1967. (A collection of papers on certain theoretical and computational aspects of nonlinear programming; Chapter 2, by Abadie, deals with the Kuhn-Tucker theorem in relation to the constraint qualification.)

Allen, R. G. D.: *Mathematical Analysis for Economists*, Macmillan & Co., Ltd., London, 1938. (A clear exposition of differential and integral calculus; determinants are discussed, but not matrices; no set theory, and no mathematical programming.)

____: *Mathematical Economics*, 2d ed., St Martin's Press, Inc., New York, 1959. (Discusses a legion of mathematical economic models; explains linear differential and difference equations and matrix algebra.)

Almon, C.: *Matrix Methods in Economics*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1967. (Matrix methods are discussed in relation to linear-equation systems, input-output models, linear programming, and nonlinear programming. Characteristic roots and characteristic vectors are also covered.)

Baumol, W. J.: *Economic Dynamics: An Introduction*, 3d ed., The Macmillan Company, New York, 1970. (Part IV gives a lucid explanation of simple difference equations; Part V treats simultaneous difference equations; differential equations are only briefly discussed.)

Braun, M.: *Differential Equations and Their Applications: An Introduction to Applied Mathematics*, 2d ed., Springer-Verlag New York Inc., 1978. (Contains interesting applications of differential equations, such as the detection of art forgeries, the spread of epidemics, the arms race, and the disposal of nuclear waste.)

Burmeister, E., and A. R. Dobell: *Mathematical Theories of Economic Growth*, The Macmillan Company, New York, 1970. (A thorough exposition of growth models of varying degrees of complexity.)

761

Coddington, E. A., and N. Levinson: *Theory of Ordinary Differential Equa-*

*tions*, McGraw-Hill Book Company, New York, 1955. (A basic mathematical text on differential equations.)

Courant, R.: *Differential and Integral Calculus* (trans. E.J. McShane), Interscience Publishers, Inc., New York, vol. I, 2d ed., 1937, vol. II, 1936. (A classic treatise on calculus.)

______, and F. John: *Introduction to Calculus and Analysis*, Interscience Publishers, Inc., New York, vol. I, 1965, vol. II, 1974. (An updated version of the preceding title.)

Dorfman, R., P. A. Samuelson, and R. M. Solow: *Linear Programming and Economic Analysis*, McGraw-Hill Book Company, New York, 1958. (A detailed treatment of linear programming, game theory, and input-output analysis.)

Franklin, J.: *Methods of Mathematical Economics: Linear and Nonlinear Programming, Fixed-Point Theorems*, Springer-Verlag New York Inc., 1980. (A delightful presentation of mathematical programming.)

Frisch, R.: *Maxima and Minima: Theory and Economic Applications* (in collaboration with A. Nataf), Rand McNally & Company, Chicago, Ill., 1966. (A thorough treatment of extremum problems, done primarily in the classical tradition.)

Goldberg, S.: *Introduction to Difference Equations*, John Wiley & Sons, Inc., New York, 1958. (With economic applications.)

Hadley, G.: *Linear Algebra*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1961. (Covers matrices, determinants, convex sets, etc.)

______: *Linear Programming*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1962. (A clearly written, mathematically oriented exposition.)

______: *Nonlinear and Dynamic Programming*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1964. (Covers nonlinear programming, stochastic programming, integer programming, and dynamic programming; computational aspects are emphasized.)

Halmos, P. R.: *Naive Set Theory*, D. Van Nostrand Company, Inc., Princeton, N.J., 1960. (An informal and hence readable introduction to the basics of set theory.)

Henderson, J.M., and R.E. Quandt: *Microeconomic Theory: A Mathematical Approach*, 3d ed., McGraw-Hill Book Company, New York, 1980. (A comprehensive mathematical treatment of microeconomic topics.)

Intriligator, M. D.: *Mathematical Optimization and Economic Theory*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1971. (A thorough discussion of optimization methods, including the classical techniques, linear and nonlinear programming, and dynamic optimization; also applications to the theories of the consumer and the firm, general equilibrium and welfare economics, and theories of growth.)

Kemeny, J.G., J.L. Snell, and G.L. Thompson: *Introduction to Finite Mathematics*, 3d ed., Prentice-Hall, Inc., Englewood Cliffs, N.J., 1974. (Covers such topics as sets, matrices, probability, and linear programming.)

Koo, D.: *Elements of Optimization: With Applications in Economics and Business*, Springer-Verlag New York Inc., 1977. (Clear discussion of classical optimization methods, mathematical programming as well as optimal control theory.)

Koopmans, T.C. (ed.): *Activity Analysis of Production and Allocation*, John Wiley & Sons, Inc., New York, 1951, reprinted by Yale University Press, 1972. (Contains a number of important papers on linear programming and activity analysis.)

762

_____: *Three Essays on the State of Economic Science*, McGraw-Hill Book Company, New York, 1957. (The first essay contains a good exposition of convex sets; the third essay discusses the interaction of *tools* and *problems* in economics.)

Leontief, W.W.: *The Structure of American Economy, 1919—1939*, 2d ed., Oxford University Press, Fair Lawn, N.J., 1951. (The pioneering work in input-output analysis.)

Samuelson, P. A.: *Foundations of Economic Analysis*, Harvard University Press, Cambridge, Mass., 1947. (A classic in mathematical economics, but very difficult to read.)

Takayama, A.: *Mathematical Economics*, The Dryden Press, Hinsdale, Ill., 1974. (Gives a comprehensive treatment of economic theory in mathematical terms, with concentration on two specific topics: competitive equilibrium and economic growth.)

Thomas, G.B., and R. L. Finney: *Calculus and Analytic Geometry*, 5th ed., Addison-Wesley Publishing Company, Inc., Reading, Mass., 1979. (A clearly written introduction to calculus.)

# 附录 IV 部分习题答案

763

## 练习 2.3

**1** (*a*) $\{x \mid x > 27\}$

**3** (*a*) $\{2,4,6,7\}$ (*c*) $\{2,6\}$ (*e*) $\{2\}$

**8** 具有 16 个子集

**9** 提示:区分符号$\not\in$和$\not\subset$。

## 练习 2.4

**1** (*a*) $\{(3,a),(3,b),(6,a),(6,b),(9,a),(9,b)\}$

**3** 否。

**5** 值域 $=\{y \mid 8 \leqslant y \leqslant 17\}$

## 练习 2.5

**2** (*a*)和(*b*)的斜率符号不同;(*a*)与(*c*)的截距不同。

**4** 允许出现负值时,还需使用第 III 象限。

**5** (*a*) $x^{19}$

**6** (*a*) $x^{6}$

## 练习 3.2

**1** $\overline{P}=3\frac{2}{9}$,和$\overline{Q}=17\frac{5}{9}$

**3** 注: In 2(*a*), $c=10$(不是 6)。

**5** 提示: $b+d=0$ 意味着 $d=-b$。

764

## 练习 3.3

**1** (*a*) $\overline{x}_1=5$,和 $\overline{x}_2=2$

3 (*a*) $\overline{P}_1=1$,和 $\overline{P}_2=-5$ (*c*) $\overline{x}_1=4, \overline{x}_2=1$, 和 $\overline{x}_3=2$

5 (*a*) $\overline{P}=1$, 和 $\overline{Q}=2$

## 练习 3.4

3 $\overline{P}_1=3\frac{6}{17}$ $\overline{P}_2=3\frac{8}{17}$ $\overline{Q}_1=11\frac{7}{17}$ $\overline{Q}_2=8\frac{7}{17}$

## 练习 3.5

1 (*b*) $\overline{Y}=(a-bd+I_0+G_0)/[1-b(1-t)]$

$\overline{T}=[d(1-b)+t(a+I_0+G_0)]/[1-b(1-t)]$

$\overline{C}=[a-bd+b(1-t)(I_0+G_0)]/[1-b(1-t)]$

3 提示：在把后两个方程代入第一个方程后，将所得到的方程视为变量 $w=Y^{\frac{1}{2}}$ 的二次方程。只有一个根 $\overline{w}_1=11$ 是可接受的，它给出 $\overline{Y}=121$ 和 $\overline{C}=91$。另一个根会导致负的 $\overline{C}$。

## 练习 4.1

1 常数列向量的元素为:0, $a$, $-c$。

## 练习 4.2

1 (*a*) $\begin{bmatrix} 4 & 2 \\ 9 & 7 \end{bmatrix}$ (*c*) $\begin{bmatrix} 12 & -3 \\ 18 & 27 \end{bmatrix}$

3 在此特定情况下，$AB$ 恰好等于 $BA=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$。

4 (*b*) $\underset{(2\times 2)}{\begin{bmatrix} 49 & 5 \\ 12 & 1 \end{bmatrix}}$ (*c*) $\underset{(2\times 1)}{\begin{bmatrix} 3x+2y \\ 4x+2y-7z \end{bmatrix}}$

5 (*a*) $x_2+x_3+x_4+x_5$ (*c*) $b(x_1+x_2+x_3+x_4)$

6 (*b*) $\sum_{i=2}^{4} a_i(x_{i+1}+i)$ (*d*) 提示：$x^0=1$ 对于 $x\neq 0$

## 练习 4.3

1 (a) $uv' = \begin{bmatrix} 15 & 5 & 45 \\ 6 & 2 & 18 \\ 9 & 3 & 27 \end{bmatrix}$ (c) $xx' = \begin{bmatrix} x_1^2 & x_1x_2 & x_1x_3 \\ x_2x_1 & x_2^2 & x_2x_3 \\ x_3x_1 & x_3x_2 & x_3^2 \end{bmatrix}$

(e) $u'v = 44$ (g) $u'u = 38$

3 (a) $\sum_{i=1}^{n} P_iQ_i$ (b) $P \cdot Q$ 或 $P'Q$ 或 $Q'P$



5 (a) $2v = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$ (c) $u - v = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$

7 (a) $d = \sqrt{27}$

9 (c) $d(v,0) = (v \cdot v)^{1/2}$

## 练习 4.4

1 (a) $\begin{bmatrix} 5 & 17 \\ 11 & 17 \end{bmatrix}$

2 否;应当为 $A - B = -B + A$。

4 (a) $k(A+B) = k[a_{ij} + b_{ij}] = [ka_{ij} + kb_{ij}] = [ka_{ij}] + [kb_{ij}] = k[a_ij] + k[b_{ij}] = kA + kB$ 你可以验证每一步骤吗?

5 提示:利用在分配律中的连续两段陈述。

## 练习 4.5

1 (a) $AI_3 = \begin{bmatrix} -1 & 8 & 7 \\ 0 & -2 & 4 \end{bmatrix}$ (c) $I_2x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

3 (a) $4 \times 3$ (c) $4 \times 1$

4 提示:自乘已知对角阵,并检验所得积矩阵的幂等条件。

## 练习 4.6

1 $A' = \begin{bmatrix} 2 & -1 \\ 4 & 3 \end{bmatrix}$ 和 $B' = \begin{bmatrix} 3 & 0 \\ 8 & 1 \end{bmatrix}$

3 提示:定义 D≡AB,并应用(4.11)。

5 提示:定义 D≡AB,并应用(4.14)。

## 练习 5.1

1 (*a*) (5.2)　(*c*) (5.3) (*e*) (5.3)

3 (*a*) 是的。　(*d*) 否。

## 练习 5.2

1 (*a*) $-6$　(*c*) 0　(*e*) $3abc-a^3-b^3-c^3$

3 $|M_b|=\begin{vmatrix} d & f \\ g & i \end{vmatrix}$　$|C_b|=-\begin{vmatrix} d & f \\ g & i \end{vmatrix}$

4 (*a*) 提示:按第 3 列展开。

5 20(不是 $-20$)



## 练习 5.3

3 (*a*) 性质 IV。(*b*) 性质 III(两行均应用)。

4 (*a*) 非奇异的。　(*c*) 奇异的。

5 (*a*) 秩 $=3$　(*c*) 秩 $<3$(具体地,$=2$)

7 $A$ 是非奇异矩阵,因为 $|A|=1-b\neq 0$。

## 练习 5.4

1 $\sum_{i=1}^{4} a_{i3}|C_{i2}|$　$\sum_{j=1}^{4} a_{2j}|C_{4j}|$

3 (*a*) 互换 $A$ 的两个对角元素;以 $-1$ 乘 $A$ 的两个非对角元素
　(*b*) 用|A|除。

4 (*a*) $E^{-1}=\frac{1}{8}\begin{bmatrix} 3 & 2 & -9 \\ -1 & 2 & -5 \\ -6 & -4 & 26 \end{bmatrix}$　(*c*) $G^{-1}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

## 练习 5.5

1 (*a*) $\bar{x}_1=5$, $\bar{x}_2=2$ (*c*) $\bar{x}_1=1$, $\bar{x}_2=2$

2 (*a*) $A^{-1}=\frac{1}{7}\begin{bmatrix} 1 & 2 \\ -2 & 3 \end{bmatrix}$;$\bar{x}=\begin{bmatrix} 5 \\ 2 \end{bmatrix}$　(*c*) $A^{-1}=\frac{1}{15}\begin{bmatrix} 1 & 7 \\ -1 & 8 \end{bmatrix}$;$\bar{x}=\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

3 (*a*) $\bar{x}_1=2, \bar{x}_2=1, \bar{x}_3=0$　(*c*) $\bar{x}=0, \bar{y}=5, \bar{z}=4$

4 提示:应用(5.8)和(5.10)。

## 练习 5.6

1 (a) $A^{-1}=\frac{1}{1-b+bt}\begin{bmatrix}1 & 1 & -b\\ b(1-t) & 1 & -b\\ t & t & 1-b\end{bmatrix}$

$$\begin{bmatrix}\overline{Y}\\ \overline{C}\\ \overline{T}\end{bmatrix}=\frac{1}{1-b+bt}\begin{bmatrix}I_0+G_0+a-bd\\ b(1-t)(I_0+G_0)+a-bd\\ t(I_0+G_0)+at+d(1-b)\end{bmatrix}$$

(b) $|A|=1-b+bt$ $\quad |A_1|=I_0+G_0-bd+a$

$|A_2|=a-bd+b(1-t)(I_0+G_0)$ $\quad |A_3|=d(1-b)+t(a+I_0+G_0)$

## 练习 5.7

1 $\overline{x}_1=69.53$，$\overline{x}_2=57.03$，$\overline{x}_3=42.58$

3 (a) $A=\begin{bmatrix}0.10 & 0.50\\ 0.60 & 0\end{bmatrix}$；矩阵方程为 $\begin{bmatrix}0.90 & -0.50\\ -0.60 & 1.00\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=\begin{bmatrix}1000\\ 2000\end{bmatrix}$。

(b) $\overline{x}_1=3333\frac{1}{3}$，$\overline{x}_2=4000$

5 $|A|=0.62$，$|A_1|=1916.16$，$|A_2|=1015.70$，和 $|A_3|=1310.47$。因 767 此，近似的：$\overline{x}_1=3090.58$，$\overline{x}_2=1638.23$，$\overline{x}_3=2113.66$（因计算时所使用的数不同，答案可能会变化）。

## 练习 6.2

1 (a) $\Delta y/\Delta x=8x+4\Delta x$ (b) $dy/dx=8x$ (c) $f'(3)=24$，$f'(4)=32$

3 (a) $\Delta y/\Delta x=5$；一个常值函数

## 练习 6.4

1 左极根 = 右极限 = 15；极限值为 15。

3 (a)5 (b) 5

## 练习 6.5

1 (a) $-3/4<x$ (c) $x<1/2$

3 ($a$) $-7<x<5$ ($c$) $-4\leqslant x\leqslant 1$

## 练习 6.6

1 ($a$)8 ($c$) 18

3 ($a$)$2\frac{1}{2}$ ($c$) 2

## 练习 6.7

2 ($a$) $N^2-7N-3$ ($b$) 是的。 ($c$) 是的。

3 ($a$) $(N+2)/(N^2+2)$ ($b$)是的。 ($c$) 定义域内连续。

6 是的，每个函数是连续、平滑的。

## 练习 7.1

1 ($a$) $dy/dx=13x^{12}$ ($c$) $dy/dx=42x^5$ ($e$) $dw/du=-2u^{-1/2}$

3 ($a$) $f'(x)=18$; $f'(1)=f'(2)=18$ ($c$) $f'(x)=10x^{-3}$; $f'(1)=10$, $f'(2)=1\frac{1}{4}$

## 练习 7.2

1 $VC=Q^3-5Q^2+14Q$; $\frac{dVC}{dQ}=3Q^2-10Q+14$ 是 MC 函数。

3 ($a$) $3(27x^2+6x-2)$ ($c$) $12x(x+1)$ ($e$) $-x(9x+14)$

4 ($b$) $MR=60-6Q$

7 ($a$) $(x^2-3)/x^2$ ($c$) $20/(x+5)^2$

8 ($a$) $a$ ($c$) $-a/(ax+b)^2$

768

## 练习 7.3

1 $-6x(5-x^2)^2$

3 ($a$) $18x(3x^2-13)^2$ ($c$) $4a(ax+b)^3$

5 $x=\frac{1}{7}y-3$, $dx/dy=\frac{1}{7}$

## 练习 7.4

1 ($a$) $\partial y/\partial x_1=6x_1^2-22x_1x_2$, $\partial y/\partial x_2=-11x_1^2+6x_2$

(c) $\partial y/\partial x_1 = 2(x_2 - 2), \partial y/\partial x_2 = 2x_1 + 3$

**3** (a) 12　　(c) 10/9

**5** (a) $U_1 = 2(x_1+2)(x_2+3)^3$，$U_2 = 3(x_1+2)^2(x_2+3)^2$

## 练习 7.5

**1** $\partial\bar{Q}/\partial a = d/(b+d) > 0$　$\partial\bar{Q}/\partial b = -d(a+c)/(b+d)^2 < 0$

$\partial\bar{Q}/\partial c = -b/(b+d) < 0$　$\partial\bar{Q}/\partial d = b(a+c)/(b+d)^2 > 0$

**2** $\partial\bar{Y}/\partial I_0 = \partial\bar{Y}/\partial\alpha = 1/(1-\beta+\beta\delta) > 0$

## 练习 7.6

**1** (a) $|J| = 0$;函数相关。

(b) $|J| = -20x_2$;函数不相关。

## 练习 8.1

**1** (a) $dy = -3(x^2+1)dx$　(c) $dy = [(1-x^2)/(x^2+1)^2]dx$

**3** (a) $dC/dY = b, C/Y = (a + b_Y)/Y$

## 练习 8.2

**1** (a) $dz = (6x+y)dx + (x - 6y^2)dy$

**2** (a) $dy = [x_2/(x_1+x_2)^2]dx_1 - [x_1/(x_1+x_2)^2]dx_2$

**3** $\varepsilon_{QP} = 2bP^2/(a + bP^2 + R^{1/2})$

**5** $\varepsilon_{XP} = -2/(Y_f^{1/2}P^2 + 1)$

## 练习 8.3

**3** (a) $dy = 3[(2x_2-1)(x_3+5)dx_1 + 2x_1(x_3+5)dx_2 + x_1(2x_2-1)dx_3]$

(c) $\partial y/\partial x_1 = 3(2x_2-1)(x_3+5)$

**4** 应用微分与全微分的定义。

## 练习 8.4

**1** (a) $dz/dy = x + 10y + 6y^2 = 10y + 9y^2$

(c) $dz/dy = -15x + 3y = 108y - 30$

769 3 $dQ/dt = [a\alpha A/K + b\beta A/L + A'(t)]K^{\alpha}L^{\beta}$

4 (*b*) $\S W/\S u = 10uf_1 + f_2 \quad \S W/\S v = 3f_1 - 12v^2 f_2$

## 练习 8.5

2 (*a*) 能确定；$dy/dx = -(3x^2 - 4xy + 3y^2)/(-2x^2 + 6xy) = -9/8$

(*b*) 能确定；$dy/dx = -(4x + 4y)/(4x - 4y^3) = 2/13$

4 在(0,0)点违背条件 $F_y \neq 0$。

5 偏导数的积等于 −1。

## 练习 8.6

1 (*c*) $(d\overline{Y}/dG_0) = 1/(S' + T' - I') > 0$

3 $(\partial\bar{P}/\partial Y_0) = D_{Y_0}/(S_{\bar{P}} - D_{\bar{P}}) > 0 \quad (\partial\bar{Q}/\partial Y_0) = D_{Y_0}S_{\bar{P}}/(S_{\bar{P}} - D_{\bar{P}}) > 0$

$(\partial\bar{P}/\partial T_0) = -S_{T_0}/(S_{\bar{P}} - D_{\bar{P}}) > 0 \quad (\partial\bar{Q}/\partial T_0) = -S_{T_0}D_{\bar{P}}/(S_{\bar{P}} - D_{\bar{P}}) < 0$

5 $(\partial\bar{Y}/\partial M_{s0}) = -(I' - S_i)/|J| > 0 \quad (\partial\bar{i}/\partial M_{s0}) = -(S_Y + M')/|J| < 0$

## 练习 9.2

1 (*a*) 当 $x = 1$ 时，$y = 11$(相对极大值)。

(*c*) 当 $x = 0$ 时，$y = 3$(相对极小值)。

2 (*a*) 临界值 $x = -1$ 位于定义域外；临界值 $x = 1$ 导致 $y = 3$(相对极小值)。

4 弹性为 1。

## 练习 9.3

1 (*a*) $f''(x) = 2a, f'''(x) = 0$ (*c*) $f''(x) = 4(1-x)^{-3}, f'''(x) = 12(1-x)^{-4}$

3 (*b*)一条直线。

5 $f(x)$上的每个点均是稳定点，但是我们知道的 $g(x)$上的唯一的稳定点是在 $x = 3$。

## 练习 9.4

1 (*a*) $f(2) = 33$ 是极大值。

(c) $f(1)=5\frac{1}{3}$是极大值；$f(5)=-5\frac{1}{3}$是极小值。

2 提示：首先按照一个变量($L$ 或 $W$)写出面积函数 $A$。

3 (d) $\overline{Q}=11$ (e) 最大利润 $=111\frac{1}{3}$。

5 (a) $k<0$ (b) $h<0$ (c) $j>0$。

7 $S$ 在产出水平 20.37(近似)达到最大。

## 练习 9.5

770

1 (a) 120 (c) 4 (e) $(n+2)(n+1)$

2 (a) $1+x+x^2+x^3+x^4$

3 (b) $-63-98x-62x^2-18x^3-2x^4+R_4$

## 练习 9.6

1 (a) $f(0)=0$ 是一个拐点。 (c) $f(0)=5$ 是相对极小值

2 (b) $f(2)=0$ 是相对极小值。

## 练习 10.1

1 (a) 是的。(b) 是的。

3 (a) $5e^{5t}$ (c) $-12e^{-2t}$

5 (a) 具有 $a=-1$ 的曲线是具有 $a=1$ 的曲线对横轴的镜像。

## 练习 10.2

1 (a) 7.388 (b) 1.649

2 (c) $1+2x+\frac{1}{2!}(2x)^2+\frac{1}{3!}(2x)^3+\ldots$

3 (a) $\$10e^{0.15}$ (b) $\$690e^{0.08}$

## 练习 10.3

1 (a) 4 (c) 4

2 (a) 2 (c) $-3$ (e) 6

3 (a) 28 (c) $\ln3-\ln B$ (f) 3

## 练习 10.4

1 要求防止函数退化为一个常值函数。

3 提示:取以 $b$ 为底的对数。

4 ($a$) $y=e^{(3\ln 8)t}$或 $y=e^{6.2385t}$ ($c$) $y=5e^{(\ln 5)t}$或 $y=5e^{1.6095t}$

5 ($a$) $t=(\ln y)/(\ln 7)$或 $t=0.5139\ln y$

($c$) $t=(3\ln 9y)/(\ln 15)$或 $t=1.1078\ln 9y$

6 ($a$) $r=\ln 1.05$ ($c$) $r=2\ln 1.03$

## 练习 10.5

1 ($a$) $2e^{2t+4}$ ($c$) $2te^{t^2+1}$ ($e$) $(2ax+b)e^{ax^2+bx+c}$

3 ($a$) $5/t$ ($c$) $1/(t+9)$ ($e$) $1/[x(1+x)]$

771 5 提示:利用(10.21),并应用链式法则。

7 ($a$) $3(8-x^2)/[(x+2)^2(x+4)^2]$

## 练习 10.6

1 $\bar{t}=1/r^2$

2 $d^2A/dt^2=-A(\ln 2)/4\sqrt{t^3}<0$

## 练习 10.7

1 ($a$) $2/t$ ($c$) $\ln b$ ($e$) $1/t-\ln 3$

3 $r_y=kr_x$

6 $|\varepsilon_d|=n$

10 $r_Q=\varepsilon_{QK}r_K+\varepsilon_{QL}r_L$

## 练习 11.2

1 $\bar{z}=3$ 极小值。

3 $\bar{z}=c$,在情况($a$)为极小值,在情况($b$)为极大值,在情况($c$)为鞍点。

5 ($a$) 任意数偶($x,y$)而非(2,3)产生了一个正的 $z$ 值。

($b$) 是。 ($c$) 否。 ($d$) 是。($d^2z=0$)。

## 练习 11.3

1 (*a*) $q = 4u^2 + 4uv + 3v^2$　(*c*) $q = 5x^2 + 6xy$

3 (*a*) 正定。　(*c*) 既非正定,也非负定。

5 (*a*) 正定。　(*c*)负定。(*e*)正定。

6 (*a*) $r_1, r_2 = \frac{1}{2}(7 \pm \sqrt{17})$;$u'Du$ 是正定的。

(*c*) $r_1, r_2 = \frac{1}{2}(5 \pm \sqrt{61})$;$u'Fu$ 是不定的。

7 $v_1 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}, v_2 = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$

## 练习 11.4

1 $\bar{z} = 0$(极小值。)

3 $\bar{z} = -11/40$(极小值。)

5 $\bar{z} = 2 - e$(极小值),是在$(\bar{x}, \bar{y}, \bar{w}) = (0,0,1)$达到的。

6 (*b*) 提示:见(11.16)。

7 (*a*) $r_1 = 2$　$r_2 = 4 + \sqrt{6}$　$r_3 = 4 - \sqrt{6}$

772

## 练习 11.5

1 (*a*) 严格凸。　(*c*) 严格凸。

2 (*a*) 严格凹　(*c*) 二者都不是。

3 否。

5 (*a*)圆盘。　(*b*) 是的。

7 (*a*)凸组合,具有 $\theta = 0.5$,　(*b*) 凸组合,具有 $\theta = 0.2$。

## 练习 11.6

1 (*a*)否。(*b*) $\overline{Q_1} = P_{10}/4$, $\overline{Q}_2 = P_{20}/4$

3 $|\varepsilon_{d1}| = 1\frac{5}{8}$　$|\varepsilon_{d2}| = 1\frac{1}{3}$　$|\varepsilon_{d3}| = 1\frac{1}{2}$

5 (*a*) $\pi = P_0 Q(a, b)(1 + \frac{1}{2} i_0)^{-2} - aP_{a0} - bP_{b0}$

## 练习 11.7

1 $(\partial \bar{a} / \partial P_{a0}) = P_0 Q_{bb} e^{-rt} / |J| < 0$　$(\partial \bar{b} / \partial P_{a0}) = -P_0 Q_{ab} e^{-rt} / |J| < 0$

2 ($a$) 4 ($b$) $(\partial \bar{a}/\partial P_0) = (Q_b Q_{ab} - Q_a Q_{bb}) P_0 (1+i_0)^{-2}/|J| > 0$

($c$) $(\partial \bar{a}/\partial i_0) = (Q_a Q_{bb} - Q_b Q_{ab}) P_0^2 (1+i_0)^{-3}/|J| < 0$

## 练习 12.2

1 ($a$) $\bar{z} = 1/2$，当 $\bar{\lambda} = 1/2$，$\bar{x} = 1$ 和 $\bar{y} = 1/2$ 时达到。

($c$) $\bar{z} = -19$，当 $\bar{\lambda} = -4$，$\bar{x} = 1$，和 $\bar{y} = 5$ 时达到。

4 $Z_\lambda = -G(x,y) = 0 \quad Z_x = f_x - \lambda G_x = 0 \quad Z_y = f_y - \lambda G_y = 0$

5 提示：区分恒等式和条件等式。

## 练习 12.3

1 ($a$) $|\bar{H}| = 4$；$\bar{z}$ 是极大值。 ($c$) $|\bar{H}| = -2$；$\bar{z}$ 是极小值。

## 练习 12.4

2 ($a$) 拟凹，但不是严格拟凹。 ($c$) 严格拟凹。

4 ($a$) 二者都不是。 ($c$) 拟凸，但不是拟凹。

5 提示：复习 9.4 节。

7 提示：使用(12.21)或(12.25′)。

## 练习 12.5

1 ($b$) $\bar{\lambda} = 3$，$\bar{x} = 16$，$\bar{y} = 11$ ($c$) $|\bar{H}| = 48$；条件满足。

3 $(\partial \bar{x}/\partial B) = 1/2P_x > 0 \quad (\partial \bar{x}/\partial P_x) = -(B + P_y)/2P_x^2 < 0$

$(\partial \bar{x}/\partial P_y) = 1/2P_x > 0$ 等。

773 5 不正确。

7 (a)与(b)均不是，参见(12.32)和(12.33′)。

## 练习 12.6

1 ($a$) 一次齐次。 ($c$) 非齐次的。

($e$) 二次齐次。

4 是的。

7 ($a$) $a + b + c$ 次齐次。

8 ($a$) $j^2Q = g(jK, jL)$ ($b$) 提示:令 $j = 1/L$。
($d$) $K$ 与 $L$ 的一次齐次。

## 练习 12.7

1 ($a$) 1:2:3 ($b$) 1:4:9
2 提示:复习图 8.2 和 8.3。
4 提示:这是一个全导数。
6 ($a$) 向下倾斜的直线。 ($b$) 当 $\rho \to -1, \sigma \to \infty$。
8 ($a$) 7 ($c$) $\ln 5 - 1$

## 练习 13.2

1 ($a$) $-8x^{-2} + c, (x \neq 0)$ ($c$) $\frac{1}{6}x^6 - \frac{3}{2}x^2 + c$
2 ($a$) $13e^x + c$ ($c$) $5e^x - 3x^{-1} + c, (x \neq 0)$
3 ($a$) $3\ln|x| + c, (x \neq 0)$ ($c$) $\ln(x^2 + 3) + c$
4 ($a$) $\frac{2}{3}(x+1)^{3/2}(x+3) - \frac{4}{15}(x+1)^{5/2} + c$

## 练习 13.3

1 ($a$) $4\frac{1}{3}$ ($b$) $3\frac{1}{4}$ ($e$) $2(\frac{a}{3} + c)$
2 ($a$) $\frac{1}{2}(e^{-2} - e^{-4})$ ($c$) $e^2(\frac{1}{2}e^4 - \frac{1}{2}e^2 + e - 1)$
3 ($b$) 待定。 ($e$) $f(x)$是黎曼积分。

## 练习 13.4

1 没有。
2 ($a$),($c$),($d$)和($e$)
3 ($a$),($c$)和($d$)收敛;($e$)发散。

## 练习 13.5

774

1 ($a$) $R(Q) = 14Q^2 - \frac{10}{3}e^{0.3Q} + \frac{10}{3}$ ($b$) $R(Q) = 10Q/(1+Q)$
3 ($a$) $K(t) = 9t^{4/3} + 25$

5 ($a$) 29,000

## 练习 13.6

1 仅考察资本,因为劳动也是生产所必需的,所以隐含的假设是 $K$ 与 $L$ 总是按固定比例使用。

3 提示:使用(6.8)。

4 提示: $\ln u - \ln v = \ln \frac{u}{v}$

## 练习 14.1

1 ($a$) $y(t) = -e^{-4t} + 3$ ($c$) $y(t) = \frac{3}{2}(1 - e^{-10t})$

3 ($a$) $y(t) = 4(1 - e^{-t})$ ($c$) $y(t) = 6e^{5t}$ ($e$) $y(t) = 8e^{7t} - 1$

## 练习 14.2

1 $D$ 曲线应更陡。

3 价格调整机制产生微分方程。

5 ($a$) $P(t) = A\exp\left(-\frac{\beta+\delta}{\eta}t\right) + \frac{\alpha}{\beta+\delta}$ ($b$) 是的

## 练习 14.3

1 $y(t) = Ae^{-5t} + 3$

3 $y(t) = e^{-t^2} + \frac{1}{2}$

5 $y(t) = e^{-6t} - \frac{1}{7}e^{t}$

6 提示:复习 13.2 节,例 17。

## 练习 14.4

1 ($a$) $y(t) = (c/t^3)^{1/2}$ ($c$) $yt + y^2t = c$

## 练习 14.5

1 ($a$) 是可分离变量方程,当写成$\frac{dy}{dt} + \frac{1}{t}y = 0$时,是线性的。

($c$) 可分离。可简化为伯努里方程。

**3** $y(t)=(A-t^2)^{1/2}$



## 练习 14.6

**1** ($a$) 向上倾斜的相位线;动态非稳定均衡。

($c$) 向下倾斜的相位线,动态稳定均衡。

**3** 导数的符号度量相位线的斜率。

## 练习 14.7

**1** $r_k = r_K - r_L$ [参见(10.25)]

**3** $r_Q = \rho + \lambda$

**5** ($a$) 绘出$(3-y)$和 $\ln y$ 作为两条分离线,然后相减。存在单一均衡(在 1 与 3 间的某个 y 值),且是动态稳定的。

## 练习 15.1

**1** ($a$) $y_p = 2/5$ ($c$) $y_p = 3$ ($e$) $y_p = 6t^2$

**3** ($a$) $y(t)=6e^t+e^{-4t}-3$ ($c$) $y(t)=e^t+te^t+3$

**6** 提示:应用洛比达法则

## 练习 15.2

**1** ($a$) $\frac{3}{2} \pm \frac{3}{2}\sqrt{3}i$ ($c$) $-\frac{1}{4} \pm \frac{3}{4}\sqrt{7}i$

**3** ($b$) 提示:当 $\theta = \pi/4$ 时,OP 为 45°线。

**5** ($a$) $\frac{d}{d\theta}\sin f(\theta) = f'(\theta)\cos f(\theta)$ ($b$) $\frac{d}{d\theta}\cos\theta^3 = -3\theta^2\sin\theta^3$

**7** ($a$) $\sqrt{3}+i$ ($c$) $1-i$

## 练习 15.3

**1** $y(t)=e^{2t}(3\cos 2t + \frac{1}{2}\sin 2t)$

**3** $y(t)=e^{-3t/2}(-\cos\frac{\sqrt{7}}{2}t + \frac{\sqrt{7}}{7}\sin\frac{\sqrt{7}}{2}t)+3$

**5** $y(t)=\frac{2}{3}\cos 3t + \sin 3t + \frac{1}{3}$

## 练习 15.4

1 (a) $P'' + \frac{m-u}{n-w}P' - \frac{\beta+\delta}{n-w} = -\frac{\alpha+\gamma}{n-w}$　$(n \neq w)$　(b) $P_p = \frac{\alpha+\gamma}{\beta+\delta}$

3 (a) $P(t) = e^{t/2}(2\cos\frac{3}{2}t + 2\sin\frac{3}{2}t) + 2$

## 练习 15.5

776

1 (a) $\frac{d\pi}{dt} + j(1-h) = j(\alpha - T - \beta U)$

(b) 无复根;不波动。

3 (c) 二者均为一阶微分方程。 (d) $h \neq 1$

4 (a) $\pi(t) = e^{-t}\left(A_5\cos\frac{\sqrt{2}}{4}t + A_6\sin\frac{\sqrt{2}}{4}t\right) + m$　(c) $\bar{p} = m; \bar{U} = \frac{1}{18} - \frac{2}{9}$
$m$

## 练习 15.6

2 (a) $y_p = t - 2$　(c) $y_p = \frac{1}{4}e^t$

## 练习 15.7

1 (a) $y_p = 4$　(c) $y_p = \frac{1}{18}t^2$

3 (a) 发散。 (c) 收敛。

## 练习 16.2

1 (a) $y_{t+1} = y_t + 7$　(c) $y_{t+1} = 3y_t - 9$

3 (a) $y_t = 10 + t$　(c) $y_t = y_0\alpha^t - \beta(1 + \alpha + \alpha^2 + \ldots + \alpha^{t-1})$

## 练习 16.3

1 (a) 非振荡;发散。　(c) 振荡;收敛。

3 (a) $y_t = -8(1/3)^t + 9$　(c) $y_t = -2(-1/4)^t + 4$

## 练习 16.4

1 $Q_t = \alpha - \beta(P_0 - \bar{P})(-\delta/\beta)^t - \beta\bar{P}$

3 (*a*) $\bar{P}=3$;放大振荡。 (*c*) $\bar{P}=2$;单位振荡。

5 供给函数滞后。

## 练习 16.5

1 $a=-1$

3 $P_t=(P_0-3)(-1.4)^t+3$,具有放大振荡。

## 练习 16.6

1 否。

2 (*b*) 非振荡,放大向下运动。

(*d*) 衰减,稳定地向 $R$ 运动。

4 (*a*) 首先向下倾斜,然后变成水平。

777

## 练习 17.1

1 (*a*) $\frac{1}{2}\pm\frac{1}{2}i$ (*c*) $\frac{1}{2}, -1$

3 (*a*) 4(稳定的) (*c*) 5(稳定的)

4 (*b*) $y_t=\sqrt{2}^t\left(2\cos\frac{\pi}{4}t+\sin\frac{\pi}{4}t\right)+1$

## 练习 17.2

1 (*a*) 子情况 1D (*c*) 子情况 1C

3 提示:使用(17.16)。

## 练习 17.3

3 可能性 *v*, *viii*, *x*, 和 *xi* 将变得可行。

4 (*a*) $p_{t+2}-[2-j(1-h)-\beta k]p_{t+1}+[1-j(1-h)-\beta k(1-j)]p_t=j\beta km$

(*c*) $\beta k \gtreqqless 4$

## 练习 17.4

1 (*a*) 1 (*c*) $3t^2+3t+1$

3 (*a*) $y_p=\frac{1}{4}t$ (*c*) $y_p=2-t+t^2$

5 (*a*) 1/2，−1 和 1

## 练习 18.2

2 $b^3+b^2-3b+2=0$

3 (*a*) $x_t=-(3)^t+4(-2)^t+7$ $y_t=2(3)^t+2(-2)^t+5$

4 (*a*) $x(t)=4e^{-2t}-3e^{-3t}+12$ $y(t)=-e^{-2t}+e^{-3t}+4$

## 练习 18.3

2 (*c*) $\beta=(\delta I-A)^{-1}u$

3 (*c*) $\beta=(\rho I+I-A)^{-1}\lambda$

5 (*c*) $x_1(t)=4e^{-4t/10}+2e^{-11t/10}+\frac{17}{6}e^{t/10}$；$x_2(t)=3e^{-4t/10}-2e^{-11t/10}+\frac{19}{6}e^{t/10}$

## 练习 18.4

3 (*a*) $$\begin{bmatrix}\pi_c\\U_c\end{bmatrix}=\begin{bmatrix}A_3e^{-t}+A_4te^{-t}+m\\A_3e^{-t}+A_4te^{-t}+\frac{1}{8}\end{bmatrix}$$

778

4 (*a*) $$\begin{bmatrix}\pi_c\\U_c\end{bmatrix}=\begin{bmatrix}A_1\\\frac{23-\sqrt{193}}{48}A_1\end{bmatrix}\left(\frac{33+\sqrt{193}}{64}\right)^t+\begin{bmatrix}A_2\\\frac{23+\sqrt{193}}{48}A_2\end{bmatrix}\left(\frac{33-\sqrt{193}}{64}\right)^t+\begin{bmatrix}m\\\frac{1-m}{6}\end{bmatrix}$$

## 练习 18.5

1 单个方程可以写成两个一阶方程。

2 是的。

4 (a) 鞍点。

## 练习 18.6

**1** (*a*) $|J_E| = 1$ 和 $\mathrm{tr}J_E = 2$;局部非稳定结点。

(*c*) $|J_E| = 5$ 和 $\mathrm{tr}J_E = -1$;局部稳定结点。

**2** (*a*) 局部鞍点。

(*c*) 局部稳定结点或焦点。

**4** (*a*) $x' = 0$ 曲线和 $y' = 0$ 曲线重合,并提供了一个均衡点的直线。

## 练习 19.1

**1** (*a*) $(\bar{x}_1, \bar{x}_2) = (2,3)$　　(*b*) $(\bar{x}_1, \bar{x}_2) = (2, \frac{1}{2})$

**3** $(\bar{x}_1, \bar{x}_2) = (0.4, 0.6)$

**5** (*a*) 目标函数将绘成一族水平线。

(*c*) 它们意味着等式约束 $ax_1 + bx_2 = c$。

**7** 无影响。

## 练习 19.2

**1** 是的。

**3** 提示:参见练习 19.1－3,(*c*)部分

## 练习 19.4

**3** (*a*) (0,0,4,3,8),(0,3,4,0,2),(2,3,2,0,0),(4,2,0,1,0),(4,0,0,3,4)

**5** (*b*),(*c*),(*d*) 和 (*e*)

## 练习 19.5

**1** $(\bar{\pi}, \bar{x}_1, \bar{x}_2, \bar{s}_1, \bar{s}_2) = (14,2,2,0,0)$

**3** $(\bar{\pi}, \bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{s}_1, \bar{s}_2, \bar{s}_3) = (34,4,0,2,0,0,5)$

## 练习 19.6



**2** $(\bar{C}, \bar{x}_1, \bar{x}_2, \bar{s}_1, \bar{s}_2) = (8,8,0,0,12)$

**4** $(\bar{C}, \bar{x}_1, \bar{x}_2, \bar{s}_1, \bar{s}_2, \bar{s}_3) = (45, 2, \frac{1}{2}, 0, 0, \frac{7}{2})$

## 练习 20.1

4 ($a$) $(\bar{y}_1, \bar{y}_2, \bar{t}_1, \bar{t}_2, \bar{t}_3) = (2,3,2,0,0)$ 和 $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{s}_1, \bar{s}_2) = (0,1,2,0,0)$

($b$) $\bar{C}^* = \bar{C} = 19$

5 ($a$) 是的。($c$) 两个。

6 ($a$) $\bar{s}_1 = 0, \bar{s}_2 = 12$, $\bar{C} = 8$ ($c$) $\bar{y}_2 = 0$ ($e$) $\bar{C}^* = \bar{C} = 8$

7 ($a$) 提示:参见练习 19.2－3。 ($b$) 不必要。 ($c$) 有效。

## 练习 20.2

1 $\pi^*$ 度量潜在消费者的数量(以千为单位),$y_1$ 表示每美元广告支出所吸引的潜在消费者数量。

## 练习 20.3

2 不能,它是劣的组合。

4 ($a$) 3－空间。 ($b$) 是的。($c$) 棱锥。

6 $(\bar{Q}_1, \bar{Q}_2, \bar{Q}_3, \bar{Q}_4) = (4,0,0,0)$。不同于 $K_0, L_0$ 不能完全就业。

## 练习 20.4

1 $\sum_{j=1}^{n} a_{0j} x_j \leqslant L_0$,或 $A'_0 x \leqslant L_0$

3 该锥面包括位于射线 $\begin{bmatrix} 1 \\ 1 \end{bmatrix} Q_1$ 其下方,以及射线 $\begin{bmatrix} 2 \\ -1 \end{bmatrix} Q_2$ 及其上方所有点。

4 ($a$) 提示:复习图 20.3 构建等量线的讨论,并将此原理应用于图 20.6b。

## 练习 21.1

1 $(\bar{x}_1, \bar{x}_2) = (5,5)$

3 $(\bar{x}_1, \bar{x}_2) = (3,0)$

5 一个垂直把持的冰激凌锥。

780

## 练习 21.2

3 条件 $x_j(\partial Z/\partial x_j) = 0$ 和条件 $y_i(\partial Z/\partial y_i) = 0$ 可以被压缩成一个方程。

5 一致。

6 (*b*) $\partial Z/\partial y_1$，$\partial Z/\partial y_2$ 和$\partial Z/\partial y_3$ 恰好均为负；因此，$\bar{y}_1=\bar{y}_2=\bar{y}_3=0$。

(d) 是的。

## 练习 21.3

1 不能求得合格的弧来作为像$(dx_1,\ dx_2)=(1,0)$这样的检验向量。

3 $(\bar{x}_1,\bar{x}_2)=(0,0)$是一个尖点。满足约束规范；(所有的检验向量均是水平的，并指向东方。)库恩—塔克条件也是满足的。

5 选择 $\bar{y}_0=0$ 和 $\bar{y}_1\geqslant 0$，所有条件都能得到满足。

## 练习 21.4

1 (*a*) 最大化 $-C=-F(x)$，满足约束 $-G^i(x)\leqslant r_i$，$(i=1,\ldots,m)$和 $x\geqslant 0$。

(c) 在非负正交分划体内，对于所有的 $i$，$F(x)$为凸，$G^i(x)$为凹。

3 以严格不等式代替弱不等式。

4 (*a*) 可以应用。

## 练习 21.5

2 充分但不是必要条件。

3 (*b*) 中的函数是可接受的，它是拟凹函数。

4 (*a*) 是的。

## 练习 21.6

1 具有 $Z=U(x_1,\ldots,x_n)+y(B-P_1x_1-\ldots-P_nx_n)$：

$\partial Z/\partial x_j=U_j-yP_j\leqslant 0 \quad \partial Z/\partial y=B-P_1x_1-\ldots-P_nx_n\geqslant 0$

加上非负性和互补松弛性条件。

3 (*b*) $R$ 凹，$C$ 凸。

(*c*) $R$ 拟凹，$C(Q)+A-R(Q,A)$拟凸。

4 (*b*) $\bar{x}_1=\bar{x}_2=1$

5 $\bar{d}=\sqrt{2}$

# 附录V 索引

784

785

787

# 译 者 后 记

蒋中一（Alphar C. Chiang）是美国著名华裔经济学家。他1946年于上海圣约翰大学获学士学位，1948年于美国科罗拉多大学获硕士学位，1954年于美国哥伦比亚大学获博士学位。1961年至1964年任丹尼森大学经济系主任，1964年起任康涅狄格大学经济系教授。他还是美国福特基金会和美国国家科学基金会成员。

近些年来，国内经济学界围绕数学在经济学中的地位和作用，展开了广泛的争论。我们不想对这些争论作任何评价。但可以肯定的是，经济学是对现实经济活动的一种抽象。用数学语言来表达经济学思想绝对不是唯一有效的方式，但它却具有显而易见的优势。首先，运用数学工具表达经济学思想，可以促使经济学家在推理的每一阶段都作出明确的假设，从而使读者对理论所赖以建立的基础、前提以及理论的适用范围有清楚的了解。其次，运用数学符号进行推理，逻辑更为严谨，表达更为简洁。第三，运用经济计量学或统计学的方法，发展、证实或者反驳某种理论的整个过程，均需建立在数理经济学的基础之上。而计算机科学的发展与大量统计及经济计量学软件包的开发，则使这种方法的作用与价值不断增大。第四，全面准确地掌握数理经济学，是我们全面深入地理解西方现代经济学思想的一个基本前提。正是基于这些原因，我把蒋中一教授的这本《数理经济学的基本方

法》翻译过来，希望能够为促进数理经济学在我国的发展产生一定的作用。

这本《数理经济学的基本方法》是蒋中一教授的代表作。该书第一版出版于1967年，并于1974年和1984年两次进行了再版。我们是根据第三版译出的。作为一本数理经济学的基础性教材，本书的最大特点在于它把经济分析循序渐进的发展过程(由静态分析、比较静态分析、最优化问题到动态分析及数学规划)与数学方法和工具由难到易、由简到繁的过程有机地结合起来，内容系统完整，通俗易懂。本书对那些学习经济学但缺乏必要的数学知识的学生来说是必不可少的，它能使他们准确地掌握数理经济学的基本方法，而又不会为深奥的数学推导所困扰。本书对那些具有一定的数学基础而又致力于学习经济学的学生也是有益的，本书对消费理论、生产理论、通货膨胀与失业模型等的介绍也有其独到之处。书中还为读者安排了例题、习题和习题解答，这对读者深入理解书中的内容，也有极大的帮助。

本书可以作为高等院校经济类各专业的数理经济学教科书。对于广大财经工作者和对经济学感兴趣的自然科学工作者而言，也是一本极好的参考书。

本书由刘学同志翻译，赵东生等同志协助翻译了19—21章的部分内容。我的导师，北京大学光华管理学院秦宛顺教授在百忙之中认真审校了全书，在此表示诚挚的谢意。

这里，我要特别感谢商务印书馆的程秋珍、张宗理女士，没有她们的支持与关怀，没有她们的辛勤工作，本书是难以在这么短的时间内问世的。我还要感谢北京大学经济学院刘文忻教授，感谢她向译者推荐了这本书。

由于译者水平有限，译文中的错误和不当之处在所难免，还望读者指正、馈教。

译　者

1998 年 6 月 16 日